

РАЗРАБОТКА ПРОГРАММНО-АППАРАТНОГО КОМПЛЕКСА НА ОСНОВЕ НЕЙРОННОЙ СЕТИ ДЛЯ ПЕРВИЧНОЙ ДИАГНОСТИКИ ГАСТРОЭНТЕРОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ

Яхшибоев Р.Э.

ТУИТ имени Мухаммада аль-Хоразмий

yaxshiboyevrustam@gmail.com

Аннотация. В данной статье рассматривается разработка программно-аппаратного комплекса для первичной диагностики гастроэнтерологических заболеваний. На базе этого программно-аппаратного комплекса был разработан алгоритм на основе нейронных сетей. Сделанный в процессе изучения набор данных, основанный на сделанном с помощью специалистов гастроэнтерологом. Научное исследование проведено совместно с клиникой Ташкентской медицинской академии в отделе гастроэнтерологии.

Ключевые слова: алгоритм, первичная диагностика, нейронная сеть, гастроэнтерология, гастроэнтерологи, аппаратно-программный комплекс.

Введение. В настоящее время развитие искусственного интеллекта во всех странах мира развивается осторожно и стремительно. Связи в развитии искусственного интеллекта Президент Республики Узбекистан Ш. Мирзиёев издал постановление «О мерах по созданию условий для ускоренного внедрения технологий искусственного интеллекта», данное постановление соответствует стратегии «Цифровой Узбекистан – 2030» [1].

В области медицины цифровые технологии могут найти широкое применение при диагностике и лечении различных заболеваний и разной степени тяжести. С помощью цифровых технологий можно облегчить работу врачей, снизить человеческий фактор, сократить время обследования и повысить эффективность.

В течение короткого времени врач может принять решение о диагнозе. С помощью цифровых технологий можно преодолеть спорные моменты. Цифровые технологии используют искусственный интеллект, нейронные сети, машинное обучение и современные языки программирования Python [8,9,10].

На основе вышеизложенного был разработан программно-аппаратный комплекс с применением современных аппаратных технологий, то есть микроконтроллеров, регуляторов, аналого-цифровых преобразователей и т.д.

На основе нейронной сети разработан алгоритм обучения набора данных для первичной диагностики гастроэнтерологических заболеваний.

В научном исследовании была взята человеческая слюна. По слюне можно предсказать гастроэнтерологические заболевания. При болезни состав слюны резко меняется. Состав слюны соответствует параметрам [10,11,12].

Изменяя состав слюны, можно создать набор данных для обучения алгоритмов искусственного интеллекта. В таблице 1 представлен состав здорового человека [13,14,15].

ТАБЛИЦА I. СОСТАВ СЛЮНЫ ЗДОРОВОГО ЧЕЛОВЕКА

№	Состав слюны	Qty. (% and g/l)
1	Вода	99.4-99.5%
2	Органические и неорганические компоненты	0.5-0.6%
3	Белки	1.4-6.4 g/l
4	Муцин	0.8-6.0 g/l
5	холестерин	0.02-0.5 g/l
6	Глюкоза	0.1-0.3 g/l
7	Аммоний	0.01-0.12 g/l
8	Мочевая кислота	0.005-0.03 g/l

II. РАЗРАБОТКА ПРОГРАММНО-АППАРАТНОГО ОБЕСПЕЧЕНИЯ

Приборы для первичной диагностики заболеваний желудочно-кишечного тракта получили название «Слюна». Устройство состоит из датчика, АЦП, микроконтроллера, UART, регулятора и Bluetooth. При разработке прибора Saliva особое внимание уделялось его дешевизне и удобству переноски.

Архитектура системы «Слюна» состоит из функциональных блоков и каждый выполняет определенные задачи, рисунок - 1. Система «Saliva» состоит из 5 основных модулей, которые составляют программно-аппаратный комплекс [2,3,4].

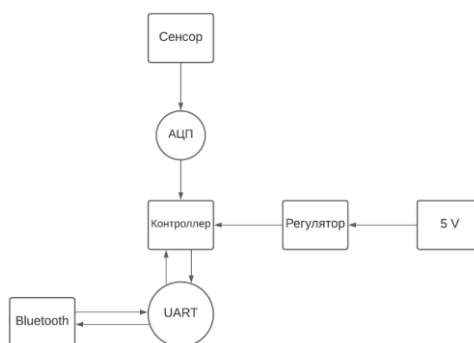


Рис 1. Функциональный модуль прибора «Saliva»

Датчики. По выбранным параметрам набора данных был собран набор сенсоров белков, глюкозы, муцина, холестерина, аммония и мочевой кислоты.

АДС1298. Аналого-цифровой преобразователь (АЦП) представляет собой маломощный многоканальный 24-разрядный дельта-сигма АЦП, разработанный компанией Texas Instruments. В микросхеме ADS1298 дискредитируются все каналы одновременно. Параметрами усиления ADS1298 можно управлять с помощью PGA. Этот чип позволяет измерять слюну пациента по выбранным параметрам, может использовать схему RLD. Передача данных чипа может достигать скорости 500-32 кбит/с (значение передаваемой дискретной частоты 103 секунды). Связь между прибором «Салива» и компьютером устанавливается по интерфейсу UART [5,6,7].

Ниже приведены технические характеристики аналого-цифрового преобразователя ADS1298:

- 8 малошумящих ПГА и 8 высокоточных АРК';
- Мощность канала: 0,75 мВт/канал.
- Уровень шума: 4 мВpp (полоса пропускания 150 Гц, G = 6).
- Ток смещения: 200 пА.
- Скорость передачи данных: 250 SPS - 32 kSPS.
- Подавление синфазного сигнала: -115 дБ.
- Программируемое усиление: 1, 2, 3, 4, 6, 8 или 12
- АЦП ААМІ ЕС11, ЕС13, ІЕС60601-1, ІЕС60601-2-27 и ІЕС60601-2-Соответствует требованиям стандарта 51.
- Обеспечивает однополярность и биполярность:
 - o AVDD = 2,7 В - 5,25 В
 - o DVDD = 1,65 В - 3,6 В
- Встроенный усилитель RLD, обнаружение срабатывания, WST терминал, определение скорости, тестовые сигналы.
- Встроенное измерение импеданса дыхания.
- Возможность измерения скорости в цифровом виде.
- Встроенный осциллятор.
- SPI-интерфейс.

Микроконтроллер Atmega328. Микроконтроллер принимает первично обработанные сигналы образцов слюны, поступающие от микросхем ADS1298 в блоке питания прибора Saliva, выполняет вторичную обработку и передает их на модуль Bluetooth по интерфейсу SPI. Микроконтроллер также управляет модулем ADS1298, 12-канальным устройством для дискретного и аналого-цифрового преобразователя сигналов пробы слюны и другими периферийными устройствами. Связь между Atmega328 и периферией реализована с помощью модуля SPI. Блок-схема микроконтроллера Atmega328 [6,7,8].

Микроконтроллер Atmega328 подключен к двум интерфейсам SPI, которые поддерживают высокоскоростную связь с устройствами ADS1298 и NS-

05. Модуль ADS1298 обеспечивает систему синхронизации последовательной связи. Синхронизация во всех коммуникационных процессах должна быть хотя бы минимальной. 8-разрядный микроконтроллер Atmega328 с низким энергопотреблением и высокой эффективностью имеет следующие особенности.

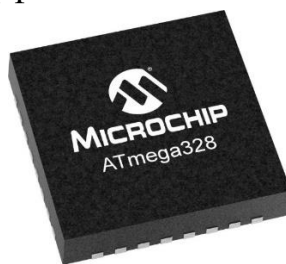


Рис 2. Микроконтроллер Atmega328.

Ниже приведены характеристики микроконтроллера Atmega328:

- Работа с совершенными короткими командами.
- AVR с 40 контактами.
- 32 КБ флэш-памяти.
- 1 КБ постоянной памяти.
- 2 КБ оперативной памяти.
- Количество входных и выходных контактов – 23.
- Таймер: 2 8-битных и 1 16-битный.
- 10-битный 6-канальный аналого-цифровой преобразователь.
- 6-канальный широкоимпульсный модулятор.
- Отдельный осциллятор.
- Поддерживает режим SPI master-slave и I²C.
- Внешний генератор 20 МГц.
- Универсальный синхронно-асинхронный приемник и передатчик.

Bluetooth HC-05. Bluetooth HC-05, один из основных модулей питания прибора «Салива», обеспечивает процесс беспроводного обмена данными между компьютером и прибором «Салива» по интерфейсу UART. Диапазон частот модуля HC-05 и канала передачи данных соответствуют ISM, т.е. 2,4 ГГц. Эта промежуточная частота представляет собой диапазон радиочастот, определенный правилами радиосвязи Международного союза электросвязи.

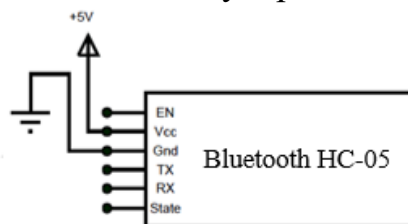


Рис. 3. Структурная схема Bluetooth-модуля HC-05.

Суммарный размер информации 8 каналов, поступающих от ADS1298, составляет $224 \times 8 = 224 \times 23 = 227$, т.е. $227 \times 50 / 8 = 838\ 860\ 800$ или $838\ 860\ 800 / 1024 = 819\ 200$ Кбит/с или 800 Мб.

Если это значение выразить в 16-й системе счисления, оно будет равно $800 / 216 = 0,0122$ Мб. Общая пропускная способность устройства Bluetooth NS05, выбранного для проектирования устройства ЭКГ, составляет 2,1 Мбит/с, что примерно в 10 раз больше, чем цифровые данные (0,0122 Мбит/с), генерируемые устройством АЦП, что соответствует текущему спросу.

Bluetooth HC-05 имеет два разных состояния (режима): состояние команды и состояние передачи/приема. 38400 кбит/с в командном режиме и 9600 кбит/с в режиме передачи/приема данных.

Ниже приведены технические характеристики Bluetooth-устройства NS05.

- Микросхема Bluetooth: HC-05 (BC417143).
- Диапазон радиочастот: 2,4–2,48 ГГц.
- Мощность передачи: 0,25–2,5 мВт.
- Чувствительность сигнала: -80 дБм (0,1% BER).
- Напряжение питания: 3,3-5В.
- Требуемый ток: 50 мА.
- Радиус поражения: до 10 метров.
- Интерфейс: последовательный порт (UART).
- Режим: ведущий, ведомый, режим ведущий/ведомый.
- Рабочая температура: -25...75 °С.
- Размеры: 27 x 13 x 2,2 мм.
- Стандартизирован на основе IEEE 802.15.1.
- FHSS – метод передачи радиосигнала, используется вид модуляции.
- Скорость передачи данных по умолчанию 38400, но 9600, 19200, 38400, 57600,
- Поддерживает 115200, 230400, 460800 бод.

Универсальный асинхронный приемопередатчик (UART). Очень простой протокол, который использует только два провода между передатчиком и приемником для передачи и приема в обоих направлениях. Оба конца также заземлены.

Связь в UART может быть симплексной (данные передаются только в одном направлении), полудуплексной (каждая сторона передает, но только по очереди) или дуплексной (обе стороны могут передавать одновременно). Данные в UART передаются в виде кадров.

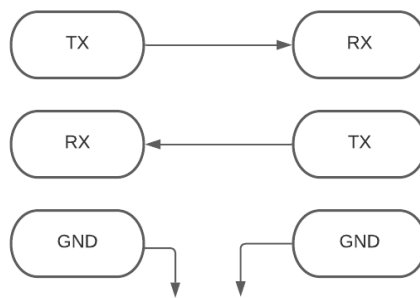


Рис 4. Данные в UART передаются в виде фреймов.

Одним из больших преимуществ UART является его асинхронность — передатчик и приемник не используют общий тактовый сигнал. Хотя это значительно упрощает протокол, это свойство накладывает определенные требования к передатчику и приемнику.

Поскольку они не используют общий тактовый сигнал, оба конца должны передавать с одной и той же заранее определенной скоростью, чтобы иметь одинаковую синхронизацию битов. Наиболее распространенные скорости передачи UART, используемые сегодня, составляют 4800, 9600, 19,2 кбит/с, 57,6 кбит/с и 115,2 кбит/с. В дополнение к одинаковой скорости передачи обе стороны соединения UART также должны использовать одинаковую структуру кадра и параметры. Лучший способ получить представление о протоколе — посмотреть на кадр UART.

Регулятор LM2596. Регуляторы серии LM2596 являются монолитными.

ИС, обеспечивающие все активные функции понижающего импульсного стабилизатора, способного управлять нагрузкой 3 А с превосходным регулированием нагрузки и линии. Эти устройства доступны с фиксированным выходным напряжением 3,3 В, 5 В, 12 В и регулируемым выходным напряжением.

Эти элементы управления, требующие минимального количества внешних компонентов, просты в использовании и включают в себя внутреннюю частотную компенсацию и осциллятор с фиксированной частотой. Серия LM2596 работает с частотой переключения 150 кГц, что позволяет использовать компоненты фильтра меньшего размера, чем требуется для низкочастотных импульсных стабилизаторов. Доступен в стандартном 5-контактном корпусе ТО-220 с несколькими различными вариантами изгиба выводов и в 5-контактном корпусе ТО-263 для поверхностного монтажа.

Разработка алгоритма на основе нейронных сетей результаты Алгоритм случайного леса («случайный лес») — алгоритм машинного обучения, предложенный Лео Брейманом и Адель Катлер, заключающийся в использовании комитета (ансамбля) деревьев решений. Случайный лес

выбирает строки и столбцы на основе дерева решений. Модели h_1, h_2, h_3, h_4 отличаются друг от друга больше, чем при использовании только мешков за счет подбора столбцов. По мере увеличения числа базовых учащихся (k) дисперсия будет уменьшаться. Когда вы уменьшаете k , дисперсия увеличивается. Но смещение остается постоянным для всего процесса. k можно найти с помощью перекрестной проверки [4]. Реализация в Scikit-learn. Для каждого дерева решений Scikit-learn вычисляет важность узла, используя важность Джини, предполагая наличие только двух дочерних узлов (бинарное дерево):

$$n_j^i = w_j C_j - w_{\text{left}(j)} C_{\text{left}(j)} - w_{\text{right}(j)} C_{\text{right}(j)}$$

- $n_i \text{ sub}(j)$ = важность узла j
- $w \text{ sub}(j)$ = взвешенное количество выборок, достигающих узла j
- $C \text{ sub}(j)$ = примесное значение узла j
- $\text{left}(j)$ = дочерний узел слева, разделенный в узле j
- $\text{right}(j)$ = дочерний узел из правого разделения в узле j

Тогда важность каждого объекта в дереве решений рассчитывается как:

$$f_i = \frac{\sum_{j:\text{node } j \text{ splits on feature } i} n_j^i}{\sum_{k \in \text{all nodes}} n_{ik}}$$

- $f_i \text{ sub}(i)$ = важность признака i
- $n_i \text{ sub}(j)$ = важность узла j

Затем их можно нормализовать до значения от 0 до 1, разделив на сумму всех значений важности признаков:

$$\text{norm}f_i = \frac{f_i}{\sum_{j \in \text{all features}} f_j}$$

Окончательная важность признака на уровне случайного леса — это среднее значение всех деревьев. Сумма значений важности объекта для каждого дерева вычисляется и делится на общее количество деревьев:

$$\text{RF}f_i = \frac{\sum_{j \in \text{all trees}} \text{norm}f_{ij}}{T}$$

- $\text{RF}f_i \text{ sub}(i)$ = важность признака, который я вычислил для всех деревьев в модели случайного леса.
- $\text{norm}f_i \text{ sub}(ij)$ = нормализованная важность признака для i в дереве j
- T = общее количество деревьев

Реализация в Spark. Для каждого дерева решений Spark вычисляет важность функции, суммируя усиление, масштабированное по количеству выборок, проходящих через узел:

$$f_i = \sum_{j; \text{nodes } j \text{ splits on features } i} S_j C_j$$

- f_i sub (i) = важность признака i
- s sub(j) = количество выборок, достигающих узла j
- C sub(j) = примесное значение узла j

Чтобы рассчитать окончательную важность функции на уровне случайного леса, сначала важность функции для каждого дерева нормализуется относительно дерева:

$$\text{norm}f_i = \frac{f_i}{\sum_{j \in \text{all features}} f_j}$$

- $\text{norm}f_i$ sub(i) = нормализованная важность объекта i
- f_i sub (i) = важность признака i

Затем значения важности объектов из каждого дерева суммируются и нормализуются:

$$\text{RF}f_i = \frac{\sum_j \text{norm } f_{ij}}{\sum_{j \in \text{all features}, k \in \text{all trees}} \text{norm}f_{ijk}}$$

- $\text{RF}f_i$ sub(i) = важность признака, который я вычислил для всех деревьев в модели случайного леса.
- $\text{norm}f_i$ sub(ij) = нормализованная важность признака для i в дереве j

Пример даты набора указан в таблице № 2. Указаны параметры и название состава слюны человека.

Таблица II. Параметры набора данных

Параметры набора данных	Название состава слюны
Параметр_1	белок
Параметр_2	Муцин
Параметр_3	холестерин
Параметр_4	Глюкоза
Параметр_5	Аммоний
Параметр_6	Мочевая кислота

Первый раз процесс обучения проводился на количестве 100 пациентов в наборе данных. (рис. 5,6)

Patient	Parameter_1	Parameter_2	Parameter_3	Parameter_4	Parameter_5	Parameter_6	
0	1	1.4	0.8	0.02	0.10	0.01	0.005
1	1	1.5	0.9	0.03	0.11	0.02	0.006
2	1	1.6	1.0	0.04	0.12	0.03	0.007
3	1	1.7	1.1	0.05	0.13	0.04	0.008
4	1	1.8	1.2	0.06	0.14	0.05	0.009
...
94	3	10.9	10.2	0.97	1.04	0.95	0.099
95	3	11.0	10.3	0.98	1.05	0.96	0.100
96	3	11.1	10.4	0.99	1.06	0.97	0.101
97	3	11.2	10.5	1.00	1.07	0.98	0.102
98	3	11.3	10.6	1.01	1.08	0.99	0.103

59 rows x 7 columns

Рис. 5. Набор данных для 100 пациентов.

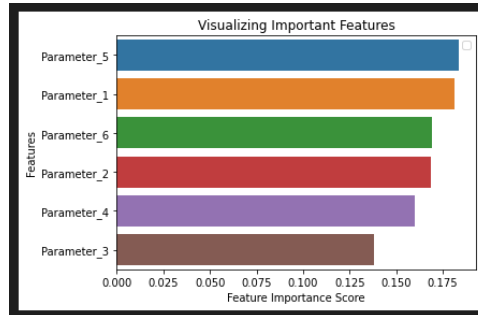


Рис. 6. Важность параметров с установленной данных (100)

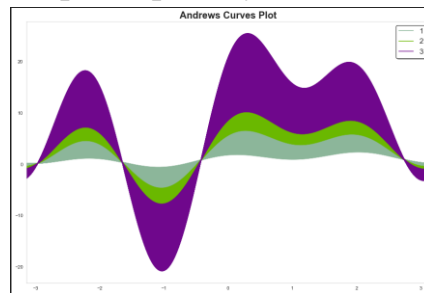


Рис. 7. Определение важности параметров и прогнозирование вероятности заболевания пациента. (Случайный лес)

- Фиолетовый - вероятность болезни выше
- Зеленый – вероятность заболевания ниже
- Фисташка – полезная

Второй раз процесс обучения проводился на количестве 1000 пациентов в наборе данных.

Patient	Parameter_1	Parameter_2	Parameter_3	Parameter_4	Parameter_5	Parameter_6	
0	1	1.40	0.80	0.020	0.10	0.010	0.0050
1	1	1.41	0.81	0.021	0.11	0.011	0.0051
2	1	1.42	0.82	0.022	0.12	0.012	0.0052
3	1	1.43	0.83	0.023	0.13	0.013	0.0053
4	1	1.44	0.84	0.024	0.14	0.014	0.0054
...
994	3	11.34	10.74	1.014	10.04	1.004	0.1044
995	3	11.35	10.75	1.015	10.05	1.005	0.1045
996	3	11.36	10.76	1.016	10.06	1.006	0.1046
997	3	11.37	10.77	1.017	10.07	1.007	0.1047
998	3	11.38	10.78	1.018	10.08	1.008	0.1048

Рис. 8. Набор данных 1000 пациентов.

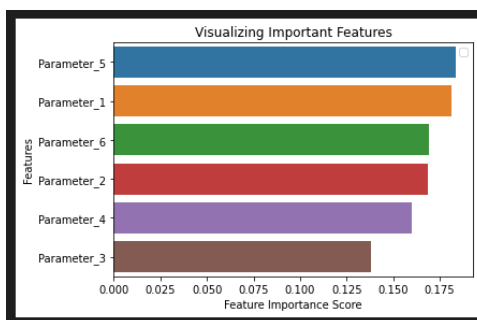


Рис 9. Важность параметров с установленной даты (1000)

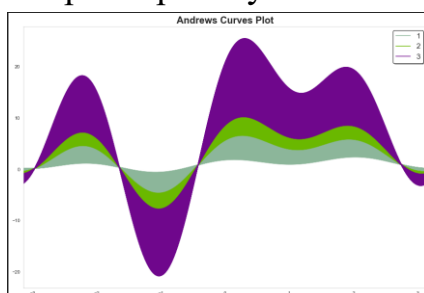


Рис 10. Определение важности параметров и прогнозирование вероятности заболевания пациента. (Случайный лес)

- Фиолетовый - вероятность болезни выше
- Зеленый – вероятность заболевания ниже
- Фисташка – полезная

Благодарность

Работа выполнена при поддержке клиники Ташкентской медицинской академии, кафедры гастроэнтерологии и кафедры биомедицинской инженерии, информатики и биофизики.

Заключение

В результате была получена 96% точность набора дат обучения от 100 до 1000 алгоритма Random Forest. В дальнейших исследованиях количество пациентов и параметры будут постепенно увеличиваться и каждый раз будут проверяться на точность.

Список использованной литературы

1. M. B. Voltaevich, N. R. H. ogli, G. N. S. qizi and M. S. S. ogli, "Estimation affects of formats and resizing process to the accuracy of convolutional neural network," *2019 International Conference on Information Science and Communications Technologies (ICISCT)*, 2019, pp. 1-5, doi: 10.1109/ICISCT47635.2019.9011858.
2. Muminov, B., et al. "Localization and Classification of Myocardial Infarction Based on Artificial Neural Network,(2020) 2020 Information Communication Technologies Conference." (2020): 245-249.

3. Ermetov E. Y. Yaxshiboyev RE Gastroenterologik kasalliklarni KNN algoritmi asosida bashoratlovchi dastur //O 'zbekiston Respublikasi intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma.№ DGU. – T. 17014.
4. Ermetov E. Y. Yaxshiboyev RE Gastroenterologik kasalliklarni ANN algoritmi asosida bashoratlovchi dastur //O 'zbekiston Respublikasi intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma.№ DGU. – T. 17016.
5. Ermetov E. Y. Yaxshiboyev RE Gastroenterologik kasalliklarni SVM algoritmi asosida bashoratlovchi dastur //O 'zbekiston Respublikasi intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma.№ DGU. – T. 17015
6. Khamzaev J. et al. DRIVER SLEEPINESS DETECTION USING CONVOLUTION NEURAL NETWORK //CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS). – 2022. – T. 1. – №. 4. – С. 31-35.
7. Яхшибоев Р. и др. РАЗРАБОТКА МОДЕЛИ РАСПОЗНАВАНИЯ ГРАФИЧЕСКИХ ОБЪЕКТОВ НА ОСНОВЕ МЕТОДА «TRANSFER LEARNING» ДЛЯ ДИАГНОСТИКИ В СФЕРЕ ЗДРАВООХРАНЕНИЯ //Conferences. – 2022. – С. 156-164.
8. Yaxshiboyev R. DEVELOPMENT OF A SOFTWARE AND HARDWARE COMPLEX FOR PRIMARY DIAGNOSTICS BASED ON DEEP MACHINE LEARNING //CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS). – 2022. – T. 1. – №. 4. – С. 20-24.
9. Яхшибоев Р., Базарбаев М., Эрметов Э. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ДИАГНОСТИКИ МЕДИЦИНСКИХ ЗАБОЛЕВАНИЙ //Conferences. – 2022. – С. 148-156.
10. Yaxshiboyev R. DEVELOPMENT OF A MODEL OF OBJECT RECOGNITION IN IMAGES BASED ON THE «TRANSFER LEARNING» METHOD //CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS). – 2022. – T. 1. – №. 4. – С. 36-41.
11. Яхшибоев Р. Э., Очилов Т. Д., Сиддиқов Б. Н. РАЗРАБОТКА ПРОГРАММНОГО СРЕДСТВА ДЛЯ ИДЕНТИФИКАЦИИ НОМЕРНЫХ ЗНАКОВ ТРАНСПОРТНЫХ СРЕДСТВ НА ОСНОВЕ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ //Journal of new century innovations. – 2022. – T. 15. – №. 1. – С. 81-93.
12. Яхшибоев Р., Сиддиқов Б. ЦИФРОВЫЕ ТЕХНОЛОГИИ ДЛЯ ПЕРВИЧНОЙ ДИАГНОСТИКИ РАЗНЫХ МЕДИЦИНСКИХ

- ЗАБОЛЕВАНИЙ //Innovations in Technology and Science Education. – 2022. – Т. 1. – №. 4. – С. 94-105.
13. Yaxshiboyev R., Yaxshiboyeva D. ANALYSIS OF ALGORITHMS FOR PREDICTION AND PRELIMINARY DIAGNOSTICS OF GASTROENTEROLOGICAL DISEASES //CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS). – 2022. – Т. 1. – №. 2. – С. 49-56.
14. Djumanov J. et al. Mathematical model and software package for calculating the balance of information flow //2021 International Conference on Information Science and Communications Technologies (ICISCT). – IEEE, 2021. – С. 1-6.
15. Yaxshiboyev R. E. et al. FORECASTING GROUNDWATER EVAPORATION USING MULTIPLE LINEAR REGRESSION //Galaxy International Interdisciplinary Research Journal. – 2021. – Т. 9. – №. 12. – С. 1101-1107.