

УДК 004.832.2

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ КВАНТОВЫХ МОДЕЛЕЙ В ПРОГНОЗЕ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ

*Адилова Ф.Т.<sup>1</sup>, + Давронов Р.Р.<sup>1</sup>*

<sup>1</sup> Институт математики им В.И. Романовского АН Республики Узбекистан,  
Ташкент, Узбекистан

+ rifqat@gmail.com

**Аннотация.** В данной работе анализируются два подхода к прогнозу сердечно-сосудистых заболеваний, которые имеют постановку задачи в форме бинарной классификации на одном и том же бенчмарке Cleveland и имеют целью выявить преимущества квантовых вычислений перед традиционными классификаторами. Это - гибридный квантовый случайный лес (HQNN, HQRF) для прогнозирования развития ишемической болезни сердца на ранних стадиях и объяснимый прогноз сердечно-сосудистых заболеваний ансамблево-квантового обучения Bagging-QSVC. В первом подходе используются различные методы выбора признаков, которые в обычно требуют больших вычислительных усилий, в то время как модель HQRF и более ранняя модель HQNN являются высокоскоростными алгоритмами в силу природы квантовых вычислений. Численные результаты показывают, что HQRF больше подходит для небольших наборов данных, в то время как HQNN лучше подходит для больших наборов данных. Во втором подходе модель Bagging-QSVC в качестве базового классификатора использует квантовый классификатор опорных векторов. Результаты работы модели объясняются через важность вклада каждого отдельного признака помощью алгоритма Shaply (SHAP). Сравнительное исследование других квантовых классификаторов, на бенчмарке Cleveland показывают превосходство Bagging-QSVC с точностью 90,16%. Из этого следует, что классификаторы квантового машинного обучения эффективнее классификаторов классического машинного обучения предсказывают заболевания сердечно-сосудистой системы.

**Ключевые слова:** машинное обучение, ансамблевое обучение, квантовое машинное обучение, объяснимое машинное обучение, прогнозирование сердечных заболеваний.

### 1 ВВЕДЕНИЕ

За последние несколько лет число смертей, вызванных болезнями сердца, значительно выросло, - в 2016 году оно составило 17,7 миллиона, но ожидается, что к 2060 году это число превысит 48 миллионов [1]. Поэтому рост смертности от сердечно-сосудистых заболеваний (ССЗ) является сегодня глобальной проблемой здравоохранения. Согласно статистике ВОЗ, за 2016 год, 82% ССЗ-пациентов проживают в странах с низким или средним уровнем дохода, а в развитых странах на ССЗ приходится 50% смертей [2]. Болезни сердца оказывают значительное влияние как на качество жизни пациентов [3], так и на экономику стран [4-6]. В Европе тратится в среднем более 200 миллиардов евро на лечение пациентов с заболеваниями сердца [5]. Поскольку ССЗ увеличивают риск некоторых других расстройств, таких как деменция [7], пневмония [8], когнитивная дисфункция [9] и болезнь Альцгеймера [10], то выявление ССЗ на ранней стадии может снизить уровень смертности, и риск других заболеваний, сокращая объем средств, расходуемых на здравоохранение [11]. Болезни сердца можно разделить на пять групп: заболевания кровеносных сосудов, нерегулярное сердцебиение, врожденный порок сердца, заболевания сердечной мышцы и клапанов сердца [12]. Из них ишемическая болезнь сердца (ИБС) является наиболее распространенным заболеванием сердца в мире [13].

Кардиология столкнулась с резким увеличением анализируемых данных, объем и качество которых даёт практикующим врачам новые возможности использовать достижения в информационных технологиях для извлечения полезной информации, улучшая тем самым процесс принятия клинических и оперативных решений [14]. Способность искусственного интеллекта (ИИ) делать выводы на основе входных данных может революционизировать оказание медицинской помощи. Машинное обучение (ML) - область искусственного интеллекта, в основном, используется в прогнозировании, диагностике, а также анализе медицинских изображений [15]. Ансамблевое обучение

(Ensemble learning, EL)- метод повышения производительности моделей ML. Успех EL -модели в области прогнозирования объясняется тем, что их прогностические характеристики превосходят характеристики составляющих ансамбль моделей.

Квантовое машинное обучение (QML), являющееся пересечением классического машинного обучения и квантовых вычислений, является новой областью ИИ, которая сегодня привлекает исследователей из самых разных дисциплин благодаря своей гибкости, адекватности представления данных, многообещающими результатами масштабируемости и скорости. Алгоритмы QML, работающие на квантовых компьютерах, способны выполнять чрезвычайно быстрые вычисления для решения задач, которые сложно, а иногда и невозможно решить на классических компьютерах. Квантовые вычисления могут обеспечить множество вариантов использования квантовых технологий, которые необходимы для продолжающейся цифровизации здравоохранения [16,17].

В данном аналитическом обзоре рассматриваются два подхода:

- гибридный квантовый случайный лес (HQRF) для прогнозирования развития ишемической болезни сердца на ранних стадиях [18];
- объяснимый прогноз сердечно-сосудистых заболеваний: подход ансамблево-квантового обучения [19].

### 1.1 Прогнозирование развития ишемической болезни сердца с помощью алгоритма гибридного квантового случайного леса

Алгоритмы машинного обучения для раннего прогнозирования ИБС, - RF, k-ближайшего соседа (KNN), NB, XGBoost, CatBoost и LightGBM и другие, широко известны, но часто имеют ограничения по требуемой вычислительной мощности и времени вычислений. Для решения этой проблемы был предложен новый алгоритм квантового машинного обучения (QML), которые могут ускорить вычисления и подходят для решения сложных задач или задач с большими объемами данных [20,21]

Перез-Салинас и др. [22] использовали метод повторной загрузки (Hot Reload) для преодоления ограничений памяти, а Сергиоли и др. [23] использовали центроидную меру Хелстрема (Helstrom Quantum Centroid, HQC) для повышения точности квантовой бинарной классификации. Автор [24] представил классификатор HQNN (hybrid quantum-classical neural network) и показал, что сеть QNN страдает от переобучения, но в гибридном формате способна ускорить обучение по сравнению с классической нейронной сетью с тем же количеством параметров. Кумар и др. [20] исследовали ряд квантовых алгоритмов: случайный лес (QRF), k-ближайший сосед (QKNN), дерево принятия решений (QDT) и гауссовский наивный байесовский алгоритм (QGNB) для выявления сердечной недостаточности. Они получили точность в 89% для набора данных Cleveland по 14 признакам. Алсубаи и др. представили метод глубокого квантового обучения для раннего прогнозирования ССЗ на данных Cleveland и достигли максимального значения AUC в 95% [25]. Наборы данных Cleveland и Statlog (часть набора Cleveland) были использованы Ауон и соавторами [26] для сравнения точности методов DNN, дерева решений, KNN, логистической регрессии, NB, RF и SVM в прогнозировании ССЗ. Их результаты показывают, что SVM с точностью 97,36% и DNN с точностью 98,15% являются наиболее точными методами для наборов данных Cleveland и Statlog соответственно. Фитриани и др. [27], проанализировали Cleveland и Statlog с использованием HDPM (Heart Disease Prediction Model), и получили точность 98,40% и 95,90% для Cleveland и Statlog соответственно. В работе [28] Амин и др. применили свой метод отбора признаков к наборам данных Cleveland и Statlog: учет девяти признаков методом голосования и использование процедур NB и логистической регрессии подходит для раннего прогнозирования ССЗ с точностью 87,41%. Основываясь на графе Lasso и формуле сжатия ковариационной матрицы Ledoit-Wolf, Карадениз и др. [29] внедрили комплексный метод выявления сердечных заболеваний на данных Statlog с точностью 85,5%. Интересный результат получен Г.Н. Ахмад, С. Уллах и др. одновременный анализ данных на базе RF из Cleveland, Hungary, Switzerland, и Long Beach V, дают результат со 100% точностью [30]. Эль-Шаффи в [31] разработал гибридный метод PSO-случайного леса для прогнозирования ССЗ и получил AUC= 92% и 91% при 10-кратной перекрестной проверке для Cleveland и Statlog соответственно.

В этом исследовании использовались хорошо известные наборы данных Cleveland и Statlog для раннего прогнозирования ишемической болезни сердца.

### 1.2 Подход ансамблево-квантового обучения

В литературе по ССЗ широко используется подход ансамблевого обучения с голосованием по большинству голосов [15,33-35]. В этом исследовании использовалась четырехэтапная методология: выбор признаков и их нормализация; анализ производительности классификаторов, -SVM и

ANN, и QML; определение эффективной QML-модели сравнением различных квантовых классификаторов, - QSVC, QNN и VQC; модель ансамблевого обучения на основе классификатора квантовых опорных векторов (Bagging-QSVC) с интерпретируемостью важности и вклада каждого признака в прогноз по методу Shaply (SHAP).

Ансамблевое обучение широко применялось для повышения точности прогнозирования сердечно-сосудистых заболеваний с использованием различных методик. Гао и др. [36] использовали методы boosting и bagging ансамблевого обучения пяти различных классических алгоритмов машинного обучения, с отбором признаков методом главных компонент (PCA), линейного дискриминантного анализа (LDA) и сравнением их показателей работы с методами ансамблевого обучения. Менве и др. [37] предложили метод ML, который случайным образом разделял набор данных на более мелкие подмножества, на которые затем строили модели с использованием дерева классификации и регрессии (Classification and Regression Tree, CART). На основе моделей CART генерировали финальную ансамблевую модель, известную как ансамбль взвешенных классификаторов старения на основе точности (Accuracy-Based Weighted Aging classifier Ensemble, AB-WAE). Лата и др. [33] на базе данных Cleveland применяли методы суммирования, бустинга, голосования большинством голосов и ансамблевого обучения. Дас и др. [38] доказали эффективность методов ансамблевого обучения в случайном подпространстве для диагностики пороков клапанов сердца. Раза [15] объединил три различных классификатора: логистическую регрессию (LR), наивный Байес (NB), искусственную нейронную сеть (ANN) и суммировал результаты этих классификаторов по большинству голосов, обеспечив лучшую производительность. Механович и др. [35] применили модель ансамблевого обучения с использованием классификаторов ANN, k-ближайшего соседа (KNN) и опорных векторов (SVM) для прогнозирования ССЗ пациентов, используя бинарную и мульти классификацию. Во всех вышеописанных случаях наибольшая точность была достигнута на ансамблевых моделях при использовании голосования большинством голосов.

Шильд и др. [39] применили ансамблевое обучение к квантовым классификаторам, которые оценивались параллельно, а доступ к их интегрированному решению осуществлялся с помощью измерения кубита. Кумар и др. [20], используя QML для выявления сердечной недостаточности в данных Cleveland, доказал, что алгоритмы машинного обучения с квантовым расширением, - K-ближайший сосед (QKNN), гауссовский наивный Байесовский алгоритм (QGNB), дерево решений (QDT) и случайный лес (QRF), дают лучшие результаты, чем традиционные алгоритмы машинного обучения. Анализируемая работа расширяет эти возможности, исследуя потенциал ансамблевого обучения на основе квантового машинного обучения в прогнозировании ССЗ и объясняя результаты, полученные с помощью модели. Рассматриваются несколько классификаторов классического и квантового машинного обучения, а также новая модель прогнозирования ССЗ для достижения точного прогнозирования, которая сочетает ансамблевое и квантовое машинное обучение. Кроме того, предлагается интерпретация модели прогнозирования ССЗ на данных Cleveland алгоритмом Shaply (SHAP).

Особенностью данного обзора является то, что оба исследования проводились на одном наборе данных, - Cleveland, который широко используется в вычислительной кардиологии.

## 2 МАТЕРИАЛЫ И МЕТОДЫ

### 2.1 Прогнозирование развития ишемической болезни сердца с помощью алгоритма гибридного квантового случайного леса

Набор данных Cleveland состоит из 303 записей с 75 признаками, 138 из которых являются здоровыми, а 165 - больными ИБС [34]. Набор данных Statlog включает 270 объектов с 14 аналогичными признаками. В данных нет пропусков, 150 объектов относятся к группе здоровых, а 120 - к группе ИБС [31].

Ставится задача бинарной классификации: класс «ИБС» и класс «Здоровые» на 14 признаках, достаточных для прогнозирования ССЗ: возраст, пол, тип боли в груди (Sp), артериальное давление в состоянии покоя (restbps), уровень холестерина (Chol), уровень сахара в крови натощак (fbs), ЭКГ-тест в состоянии покоя (Rest ekg), максимальная достигнутая частота сердечных сокращений (Thalch), стенокардия, вызванная физической нагрузкой (Exang), депрессия сегмента ST, вызванная физической нагрузкой по сравнению с состоянием покоя, наклон сегмента ST (Slope), количество крупных сосудов (Ca), Результаты стресс-теста таллием (Thai), выход модели (ИБС/нормальный).

В качестве основного метода решения предлагается использовать гибридные квантово-классические модели, которые могут значительно повысить скорость вычислений и снизить вычислительные затраты [32]. Конкретно, - гибридный классико-квантовый алгоритм бинарной классификации, основанный на случайных лесах и квантовых нейронных сетях (HQRF). Используя алгоритм QNN

[24], уменьшают число входных параметров и устраняют проблему переобучения. Из 14 первые 13 элементов являются входными, а последний - выходным. Алгоритм был независимо запущен пять раз, и каждый раз признаки распределялись случайным образом.

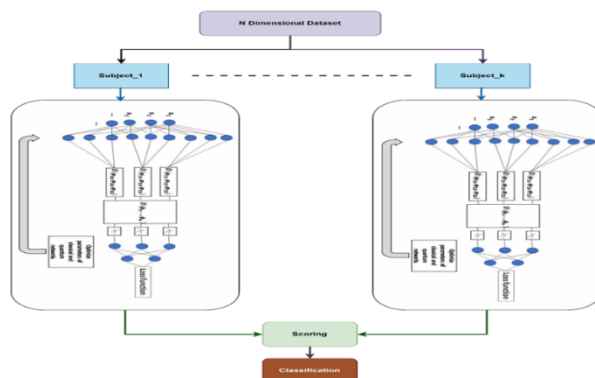


Рис. 1. Схема HQRF [18]

## 2.2 Подход ансамблево-квантового обучения

Исследование было разделено на четыре отдельных этапа, первый из которых касался набора данных Cleveland с использованием различных методов предварительной обработки. На втором этапе классические классификаторы (SVC и ANN) сравнивались с квантовыми классификаторами. На третьем этапе были исследованы три разных классификатора квантового машинного обучения (SVC, QNN и VQC). Наконец, на четвертом этапе была разработана модель ансамблевого обучения на основе квантового классификатора опорных векторов (Bagging-SVC).

Набор данных здесь был тот же самый, что и в первом подходе, - Cleveland, который состоит из 303 объектов и 14 признаков. Априори данные были размечены на 2 класса: класс больных ССЗ (метка 1), и класс здоровых (метка 0). 165 записей имеют метку 1, а 138 экземпляров имеют метку 0, т.е. сильный дисбаланс классов нет.

В квантовых компьютерах метод опорных векторов SVC является квантовым аналогом классического SVC. Используя квантовое ядро в алгоритмах QSVC, квантовые компьютеры могут ускорить обучение в SVC. В алгоритме QSVC был использован классико-квантовый метод (CQ), который предполагает использование классического набора данных на квантовых компьютерах. Классические векторы признаков преобразуют в квантовые пространства с использованием 5-кубитной карты признаков (число кубитов в схеме равно числу выбранных признаков: количеству крупных сосудов, типам боли в груди, сердечному тесту с использованием таллия, стенокардии, вызванной физической нагрузкой, и наклону сегмента ST). После подгонки классификатора QSVC к обучающим данным и оценки производительности модели с использованием тестовых данных для каждого классического входного сигнала двоичные измерения декодируют квантовые данные в соответствующие выходные данные (значение 0 или 1).

Модель Bagging-QSVC, разработанная на четвертом этапе, состоит из ансамблевой модели с классификатором QSVC. Модель QSVC была выбрана, поскольку она обеспечивает самую высокую производительность среди трех проверенных квантовых классификаторов. Тем не менее, в методе bagging ensemble каждая модель была обучена на случайной выборке из набора данных, где случайные выборки имеют тот же размер, что и исходный набор данных. Bagging-модели используют выборку с заменой, которая дублирует или игнорирует некоторые объекты из исходного набора данных в каждой выборке bootstrap. Поэтому каждая подвыборка имела разные объекты, и эти подвыборки использовались для параллельного обучения 100 различных моделей QSVC. После того как отдельные модели в ансамбле были обучены, ансамбль объединил их прогнозы, определив класс, набравший большинство голосов, чтобы получить окончательный результат модели ансамбля. Предлагаемая модель Bagging-QSVC проиллюстрирована на рис. 3.

## 3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

### 3.1 Прогнозирование развития ишемической болезни сердца с помощью алгоритма гибридного квантового случайного леса

Авторы сравнили производительность предложенного алгоритма HQRF с более ранним алгоритмом гибридной нейронной сети (HQNN) для разного числа кубитов и слоев (L) [24].

Для этого используются 10-кратная перекрестная проверка и соотношение «обучение -тест» 70/30 при разделении исходных данных, как показано в таблицах 1 и 2.

**Таблица 1.** Сравнение AUC в наборе данных Cleveland при 10-кратной перекрестной проверке [18]

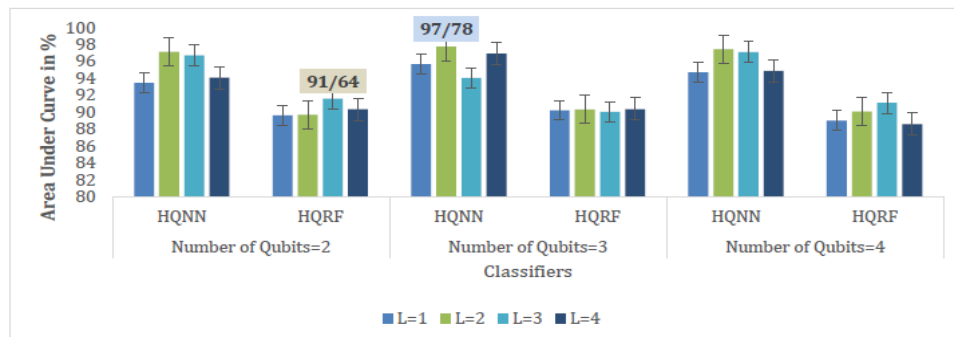
Количество Qubit = 4				Количество Qubit = 3				Количество Qubit = 2			
L=4	L=3	L=2	L=1	L=4	L=3	L=2	L=1	L=4	L=3	L=2	L=1
<b>HQNN</b>											
95.59	94.5	94.2	92.03	93.72	<b>96.43</b>	95.59	93.05	88.47	92.61	91.78	91.73
<b>HQRF</b>											
88.74	90.36	88.91	90.37	89.34	90.27	<b>91.14</b>	90.15	88.1	89.34	90.12	88

**Таблица 2.** Площадь под кривой (AUC) для HQNN и HQRF с коэффициентом разделения 70/30 (тренировка/тест) в Cleveland при использовании разного количества кубитов и слоев [18]

Количество Qubit = 4				Количество Qubit = 3				Количество Qubit = 2			
L=4	L=3	L=2	L=1	L=4	L=3	L=2	L=1	L=4	L=3	L=2	L=1
<b>HQNN</b>											
95.92/ 90.67	94.50/ 89.06	94.77/ 90.18	<b>94.89/ 93.78</b>	96.98/ 92.75	94.92/ 90.13	93.98/ 91.88	94.26/ 93.34	93.01/ 92.22	94.13/ 90.76	93.45/ 88.09	93.30/ 93.24
<b>HQRF</b>											
95.85/ 92.8	94.71/ 92.85	91.93/ 93.29	<b>89.59/ 94.36</b>	92.23/ 91.44	92.31/ 92.12	90.64/ 93.92	85.44/ 85.47	85.02/ 85.64	85.79/ 89.06	93.61/ 81.58	86.69/ 94.31

Из анализа таблиц 1 и 2 следует, что HQRF превосходит HQNN при 10-кратной перекрестной валидации в соотношении 70/30 между обучением и тестированием. Таким образом, HQNN требует большего набора обучающих данных, в то время как HQRF больше подходит для небольшого набора данных. На наборе данных Statlog были проведены аналогичные вычислительные эксперименты.

На рисунке 2 показана площадь под кривой (AUC) для моделей HQNN и HQRF в наборе данных Statlog с 10-кратной перекрестной проверкой. Максимальная средняя AUC, составляющая 97,78%, достигается при применении HQNN (3 кубита) и 91,64% - при применении HQRF (2 кубита). Результаты для соотношения "обучение/тест" 70/30 при различном количестве кубитов и слоев приведены в таблице 3.



**Рис. 2.** Площадь под кривой (AUC) для HQNN и HQRF с 10-кратной перекрестной проверкой для набора данных Statlog с использованием разного количества кубитов и слоев [18]

**Таблица 3.** Площадь под кривой (AUC) для HQNN и HQRF с соотношением 70/30 между обучением и тестированием в наборе данных Statlog с использованием различного количества кубитов и слоев [18]

Количество Qubit = 4				Количество Qubit = 3				Количество Qubit = 2			
L=4	L=3	L=2	L=1	L=4	L=3	L=2	L=1	L=4	L=3	L=2	L=1
<b>HQNN</b>											
99.17/ 82.65	99.21/ 82.96	98.76/ 88.14	99.22/ 84.87	98.15/ 79.16	98.40/ 82.83	99.00/ 76.29	98.83/ 80.61	98.83/ 76.72	98.43/ 79.93	99.05/ 84.19	97.97/ 88.33
<b>HQRF</b>											
96.55/ 86.17	95.07/ 88.76	95.47/ 85.67	89.25/ 89.93	95.52/ 86.23	92.75/ 81.97	91.56/ 81.72	87.60/ 90.52	86.99/ 76.29	88.21/ 74.62	85.15/ 80.21	86.97/ 76.35

При разделении данных в пропорции «обучение/тест» 70/30 метод HQRF с тремя кубитами и одним слоем показал наилучшие результаты. Целесообразно использовать HQRF для соотношения "обучение/тест" 70/30, а HQNN - для 10-кратной перекрестной проверки. Таким образом, HQNN

является лучшим методом для больших наборов данных, в то время как HQRF лучше подходит для небольших наборов данных, и при этом HQRF может обнаруживать отклонения в данных, и игнорировать их при классификации.

Количество кубитов в квантовых вычислениях является критическим параметром, и HQRF использует меньшее количество кубитов на каждом этапе моделирования, что делает данные небольших размеров более подходящими для квантового моделирования.

Для оценки гибридных квантовых алгоритмов HQNN и HQRF использовали наборы кардиологических данных Cleveland и Statlog. Более ранние исследования проводились с использованием различных методов выбора признаков, которые требуют больших вычислительных усилий и отнимают много времени, в то время как HQNN и HQRF являются высокоскоростными алгоритмами в силу природы квантовых вычислений.

Численные результаты показывают, что HQRF больше подходит для небольших наборов данных, в то время как HQNN лучше подходит для больших наборов данных. Кроме того, в отличие от HQRF, HQNN обладает высокой чувствительностью к выбросам.

### 3.2 Подход ансамблево-квантового обучения

Для оценки эффективности прогностических моделей, примененных в этой работе (QSVC, SVC, QNN, ANN, VQC и Bagging-QSVC), был использован набор показателей эффективности, включая точность, специфичность, чувствительность, показатель F1 и площадь под кривой (AUC) или ROC-индекс.

В табл. 4 приведены результаты, основанные на вышеупомянутых показателях эффективности классического и квантовых классификаторов, а также модели (Bagging-QSVC). Согласно результатам эксперимента, QSVC достиг точности 88,52%, в то время как традиционный SVC достиг 85,24%. Аналогичным образом, QNN достигла точности в 86,84%, в то время как ANN достигла только 85,24%. QSVC, QNN и VQC достигли наивысшей производительности для квантовых классификаторов - 88,52%, 86,84% и 85,25% соответственно. Это указывает на то, что квантовые классификаторы способны давать лучшие результаты, чем их аналоги - классические классификаторы. Bagging-QSVC по сравнению с другими квантовыми и классическими классификаторами за счет использования ансамблевого обучения, превзошел все другие классификаторы с точностью 90,16%, что указывает на то, что ансамблевое обучение улучшает производительность квантовых классификаторов.

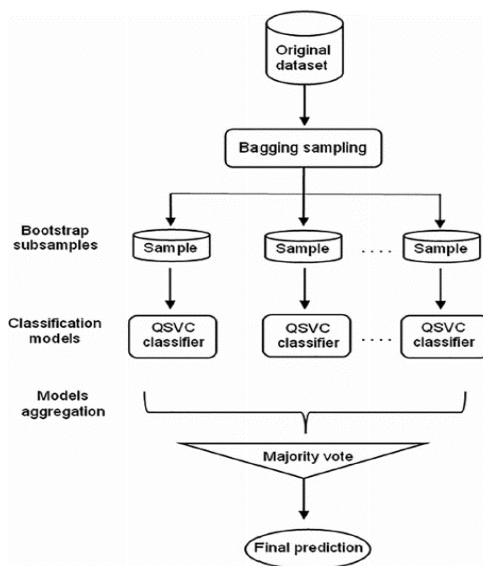


Рис. 3. Архитектура Bagging-QSVC [19]

Таблица 4. Результаты работы прогностических моделей [19]

Классификатор	Точность	Специфичность	Чувствительность	F1
SVM	85.24%	0.85	0.85	0.85
ANN	85.24%	0.85	0.85	0.85
QSVC	88.52%	0.88	0.89	0.89
QNN	86.84%	0.88	0.86	0.87
VQC	86.89%	0.85	0.85	0.85
Bagging-QSVC	90.16%	0.90	0.90	0.90

Для интерпретации результатов моделирования был использован фреймворк SHAP [40], в котором для вычисления и визуализации значимости каждого признака в прогнозе была использована библиотека SHAP на языке python. Значение SHAP - показатель вклада признака, который используется для улучшения интерпретируемости моделей машинного обучения.

График Beeswarm представляет значения признаков в порядке убывания их важности в прогнозе, как положительной, так и отрицательной. Значения SHAP используются для отображения влияния более высоких и более низких значений признаков на выходные данные модели. Каждая точка на графике представляет собой отдельное наблюдение, а цвет показывает, как более высокие и более низкие значения признака влияют на результат, при этом красный цвет обозначает более высокое значение, а синий - более низкое значение признака. На рис. 4 показан график Beeswarm, из которого видно, что наибольшее влияние на прогноз оказывает количество крупных сосудов, потом сканирование сердца с помощью таллия. Стенокардия, вызванная физической нагрузкой, оказывает наименьшее влияние на прогноз вероятности сердечных заболеваний (наименьшее значение). Значения отклонения ST указывают на то, что восходящее и плоское отклонение ST связано с более низким риском сердечных заболеваний, тогда как нисходящее отклонение ST связано с более высоким риском сердечных заболеваний.

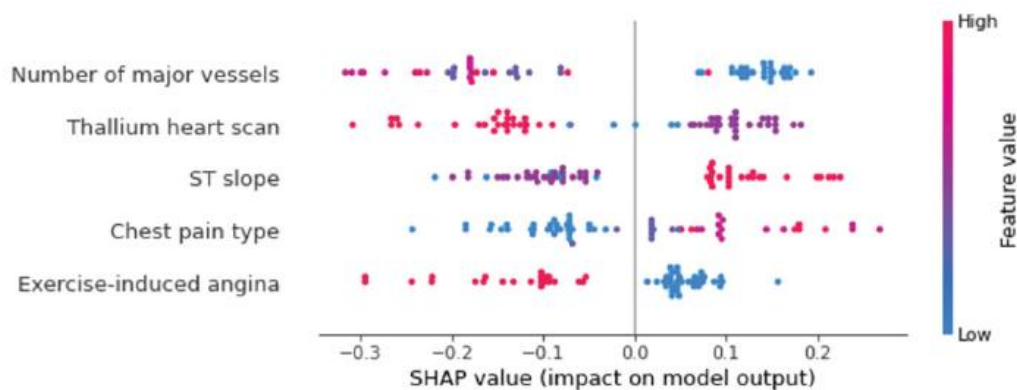


Рис. 4. График: SHAP Beeswarm [19]

Модель Bagging-QSVC предполагает случайное разбиение набора данных на более мелкие подмножества и моделирование каждого подмножества с использованием QSVC. На этой основе затем был сформирован ансамбль с использованием метода голосования большинством голосов. В наборе данных Cleveland предложенная ансамблевая модель достигла более высокой точности классификации - 90,16% по сравнению с 88,52%, 86,84% и 85,25% для QSVC, QNN и VQC, соответственно, и 85,24% как для SVC, так и для ANN. Соответственно, различные показатели эффективности и ROC-кривые показали, что предложенная модель работает лучше, чем другие модели машинного обучения. Таким образом, получено ещё одно доказательство того, что квантовые классификаторы более эффективны, чем классические, а модели ансамбля повышают производительность квантовых классификаторов в диагностике ССЗ.

#### 4 ЗАКЛЮЧЕНИЕ

Для оценки гибридных квантовых алгоритмов HQNN и HQRF использовали наборы кардиологических данных Cleveland и Statlog. Более ранние исследования проводились с использованием различных методов выбора признаков, которые требуют больших вычислительных усилий и отнимают много времени, в то время как HQNN и HQRF являются высокоскоростными алгоритмами в силу природы квантовых вычислений. Численные результаты показывают, что HQRF больше подходит для небольших наборов данных, в то время как HQNN лучше подходит для больших наборов данных. Кроме того, в отличие от HQRF, HQNN обладает высокой чувствительностью к выбросам.

Модель Bagging-QSVC предполагает случайное разбиение набора данных на более мелкие подмножества и моделирование каждого подмножества с использованием QSVC. На этой основе формируется ансамбль с использованием метода голосования большинством голосов. В наборе данных Cleveland предложенная ансамблевая модель достигла более высокой точности классификации - 90,16% по сравнению с 88,52%, 86,84% и 85,25% для QSVC, QNN и VQC, и 85,24% как для SVC, так и для ANN. Различные показатели эффективности и ROC-кривые показали, что предложенная модель работает лучше, чем другие модели машинного обучения.

Таким образом, данная статья приводит два достоверных доказательства того, что квантовые классификаторы более эффективны, чем классические, а модели ансамбля ещё и повышают производительность квантовых классификаторов в диагностике сердечно-сосудистых заболеваний.

## ЛИТЕРАТУРА

- [1] K.E. Sleeman, M. de Brito, S. Etkind, K. Nkhoma, P. Guo, I.J. Higginson, B. Gomes, R. Harding, The escalating global burden of serious health-related suffering: projections to 2060 by world regions, age groups, and health conditions, *The Lancet Global Health*. 7 (2019) e883–e892. [https://doi.org/10.1016/S2214-109X\(19\)30172-X](https://doi.org/10.1016/S2214-109X(19)30172-X).
- [2] X.-Y. Gao, A. Amin Ali, H. Shaban Hassan, E.M. Anwar. Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method, *Complexity*. 2021 (2021) 1–10. <https://doi.org/10.1155/2021/6663455>.
- [3] D.R. Thompson, C.F. Ski, J. Garside, F. Astin, A review of health-related quality of life patient-reported outcome measures in cardiovascular nursing, *European Journal of Cardiovascular Nursing*. 15 (2016) 114–125. <https://doi.org/10.1177/1474515116637980>.
- [4] S. Saha, U.-G. Gerdtham, P. Johansson, Economic Evaluation of Lifestyle Interventions for Preventing Diabetes and Cardiovascular Diseases, *IJERPH*. 7 (2010) 3150–3195. <https://doi.org/10.3390/ijerph7083150>.
- [5] A. Timmis, P. Vardas, N. Townsend, A. Torbica, H. Katus, D. De Smedt, C.P. Gale, A.P. Maggioni, S.E. Petersen, R. Huculeci, D. Kazakiewicz, V. De Benito Rubio, B. Ignatiuk, Z. Raisi-Estabragh, A. Pawlak, E. Karagiannidis, R. Treskes, D. Gaita, J.F. Beltrame, A. McConnachie, I. Bardinet, I. Graham, M. Flather, P. Elliott, E.A. Mossialos, F. Weidinger, S. Achenbach, European Society of Cardiology: cardiovascular disease statistics 2021, *European Heart Journal*. 43 (2022) 716–799. <https://doi.org/10.1093/eurheartj/ehab892>.
- [6] N.B. Oldridge, Economic burden of physical inactivity: healthcare costs associated with cardiovascular disease, *European Journal of Cardiovascular Prevention & Rehabilitation*. 15 (2008) 130–139. <https://doi.org/10.1097/HJR.0b013e3282f19d42>.
- [7] M.L. Adams, J. Grandpre, D.L. Katz, D. Shenson, Cognitive Impairment and Cardiovascular Disease: A Comparison of Risk Factors, Disability, Quality of Life, and Access to Health Care, *Public Health Rep*. 135 (2020) 132–140. <https://doi.org/10.1177/0033354919893030>.
- [8] C. Melenotte, A. Silvin, A.-G. Goubet, I. Lahmar, A. Dubuisson, A. Zumla, D. Raoult, M. Merad, B. Gachot, C. Hénon, E. Solary, M. Fontenay, F. André, M. Maeurer, G. Ippolito, M. Piacentini, F.-S. Wang, F. Ginhoux, A. Marabelle, G. Kroemer, L. Derosa, L. Zitvogel, Immune responses during COVID-19 infection, *OncoImmunology*. 9 (2020) 1807836. <https://doi.org/10.1080/2162402X.2020.1807836>.
- [9] A. Krishnaswami, M.A. Steinman, P. Goyal, A.R. Zullo, T.S. Anderson, K.K. Birtcher, S.J. Goodlin, M.S. Maurer, K.P. Alexander, M.W. Rich, J. Tjia, Deprescribing in Older Adults With Cardiovascular Disease, *Journal of the American College of Cardiology*. 73 (2019) 2584–2595. <https://doi.org/10.1016/j.jacc.2019.03.467>.
- [10] C.R. González, M. López, AMP-activated protein kinase: ‘a cup of tea’ against cholesterol-induced neurotoxicity, *J. Pathol*. 222 (2010) 329–334. <https://doi.org/10.1002/path.2778>.
- [11] A. Mehmood, M. Iqbal, Z. Mehmood, A. Irtaza, M. Nawaz, T. Nazir, M. Masood, Prediction of Heart Disease Using Deep Convolutional Neural Networks, *Arabian Journal for Science and Engineering*. 46 (2021) 3409–3422. <https://doi.org/10.1007/s13369-020-05105-1>.
- [12] S. Kalayinia, F. Arjmand, M. Maleki, M. Malakootian, C.P. Singh, MicroRNAs: roles in cardiovascular development and disease, *Cardiovascular Pathology*. 50 (2021) 107296. <https://doi.org/10.1016/j.carpath.2020.107296>.
- [13] B. Simmons, Investigating Heart Disease Datasets and Building Predictive Models, Elizabeth City State University, 2021. [https://libres.uncg.edu/ir/ecsuf/Brandon\\_Simmons\\_Thesis-Final.pdf](https://libres.uncg.edu/ir/ecsuf/Brandon_Simmons_Thesis-Final.pdf).
- [14] EIT Health and McKinsey & Company, “Transforming healthcare with AI,” 2020. [Online]. Available [https://eithealth.eu/wp-content/uploads/2020/03/EIT-Health-and-McKinsey\\_Transforming-Healthcare-with-AI.pdf](https://eithealth.eu/wp-content/uploads/2020/03/EIT-Health-and-McKinsey_Transforming-Healthcare-with-AI.pdf).
- [15] K. Raza, Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule. In: *U-Healthcare Monitoring Systems*, 1st. ed., vol. 1. New Delhi, India: Elsevier Inc, pp. 179–196, 2019.

- [16] IBM Institute for Business Value, “Exploring computing quantum use cases for healthcare,” 2020. <https://www.ibm.com/thought-leadership/institute-business-value/report/exploring-quantum-financial#>.
- [17] *Hanif Heidari, Gerhard Hellstern, Murugappan Murugappan* Heart Disease Detection using Quantum Computing and Partitioned Random Forest Methods arXiv:2208.08882v3 [quant-ph] <https://doi.org/10.48550/arXiv.2208.08882>.
- [18] *Ghada Abdulsalam, Souham Meshoul, and Hadil Shaiba* Explainable Heart Disease Prediction Using Ensemble-Quantum Machine Learning Approach IASC, 2023, vol.36, no.1 Intelligent Automation & Soft Computing DOI: 10.32604/iasc.2023.032262.
- [19] *Y. Kumar, A. Koul, P. S. Sisodia, J. Shafi, K. Verma et al.*, “Heart failure detection using quantum-enhanced machine learning and traditional machine learning techniques for internet of artificially intelligent medical things,” *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, pp. 1–16, 2021.
- [20] *R. Narain, S. Saxena, A. Kumar*, Cardiovascular Disease Prediction based on Physical Factors using Quantum Neural Network, 8 (2014).
- [21] *A. Pérez-Salinas, Alba Cervera-Lierta, E. Gil-Fuster, J.I. Latorre*, Data re-uploading for a universal quantum classifier, *Quantum*. 4 (2020) 226. <https://doi.org/10.22331/q-2020-02-06-226>.
- [22] *G. Sergioli, R. Giuntini, H. Freytes*, A new quantum approach to binary classification, *PLoS ONE*. 14 (2019) e0216224. <https://doi.org/10.1371/journal.pone.0216224>.
- [23] *G. Hellstern*, Analysis of a hybrid quantum network for classification tasks, *IET Quantum Communication*. 2 (2021) 153–159. <https://doi.org/10.1049/qt2.12017>.
- [24] *S. Alsubai, A. Alqahtani, A. Binbusayyis, M. Sha, A. Gumaei, S. Wang*, Heart Failure Detection Using Instance Quantum Circuit Approach and Traditional Predictive Analysis, *Mathematics*. 11 (2023) 1467. <https://doi.org/10.3390/math11061467>.
- [25] *S.I. Ayon, M.M. Islam, M.R. Hossain*, Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques, *IETE Journal of Research*. 0 (2020) 1–20. <https://doi.org/10.1080/03772063.2020.1713916>.
- [26] *N.L. Fitriyani, M. Syafrudin, G. Alfian, J. Rhee*, HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System, *IEEE Access*. 8 (2020) 133034–133050. <https://doi.org/10.1109/ACCESS.2020.3010511>.
- [27] *M.S. Amin, Y.K. Chiam, K.D. Varathan*, Identification of significant features and data mining techniques in predicting heart disease, *Telematics and Informatics*. 36 (2019) 82–93. <https://doi.org/10.1016/j.tele.2018.11.007>.
- [28] *T. Karadeniz, G. Tokdemir, H.H. Maraş*, Ensemble Methods for Heart Disease Prediction, *New Generation Computing*. 39(2021)569–581. <https://doi.org/10.1007/s00354-021-00124-4>.
- [29] *G.N. Ahmad, S. Ullah, A. Algethami, H. Fatima, S.Md.H. Akhter*, Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique With and Without Sequential Feature Selection, *IEEE Access*. 10 (2022) 23808–23828. <https://doi.org/10.1109/ACCESS.2022.3153047>.
- [30] *M.G. El-Shafiey, A. Hagag, E.-S.A. El-Dahshan, M.A. Ismail*, A hybrid GA and PSO optimized approach for heart-disease prediction based on random forest, *Multimed Tools Appl*. 81 (2022) 18155–18179. <https://doi.org/10.1007/s11042-022-12425-x>.
- [31] *A. Callison, N. Chancellor*, Hybrid quantum-classical algorithms in the noisy intermediate-scale quantum era and beyond, *Phys. Rev. A*. 106 (2022) 010101. <https://doi.org/10.1103/PhysRevA.106.010101>.
- [32] *C. B. C. Latha and S. C. Jeeva*, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,” *Informatics in Medicine Unlocked*, vol. 16, pp. 1–9, 2019.
- [33] *M. N. Uddin and R. K. Halder*, “An ensemble method based multilayer dynamic system to predict cardiovascular disease using machine learning approach,” *Informatics in Medicine Unlocked*, vol. 24, no. 7, pp. 1–19, 2021.
- [34] *D. Mehanović, Z. Mašetić and D. Kečo*, “Prediction of heart diseases using majority voting ensemble method,” in *IFMBE Proc.*, Banja Luka, Bosnia & Herzegovina, pp. 491–498, 2020.
- [35] *X. Gao, A. A. Ali, H. S. Hassan and E. M. Anwar*, “Improving the accuracy for analyzing heart diseases prediction based on the ensemble method,” *Complexity*, vol. 2021, pp. 1–10, 2021.
- [36] *I. D. Mienye, Y. Sun and Z. Wang*, “An improved ensemble learning approach for the prediction of heart disease risk,” *Informatics in Medicine Unlocked*, vol. 20, no. 8, pp. 1–5, 2020.

- [37] R. Das and A. Sengur, "Evaluation of ensemble methods for diagnosing of valvular heart disease," *Expert Systems with Applications*, vol. 37, no. 7, pp. 5110–5115, 2010.
- [38] M. Schuld and F. Petruccione, "Quantum ensembles of quantum classifiers," *Scientific Reports*, vol. 8, no. 1, pp.1–12, 2018.
- [39] S. Lundberg and S. Lee, "A Unified approach to interpreting model predictions," in 31st Conf. on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, pp. 1–10, 2017.

Поступила в редакцию 10.03.2025

**Цитирование:** Адилова Ф.Т., Давронов Р.Р. (2025). Сравнительный анализ квантовых моделей в прогнозе сердечно-сосудистых заболеваний. *Международный журнал теоретических и прикладных вопросов цифровых технологий*, 8(2), –С. 7-16. <https://doi.org/10.62132/ijdt.v8i2.259>.

## COMPARATIVE ANALYSIS OF QUANTUM MODELS IN THE PROGNOSIS OF CARDIOVASCULAR DISEASES

Adilova F.T.<sup>1</sup>, + Davronov R.R.<sup>1</sup>

<sup>1</sup> V.I. Romanovsky Institute of Mathematics of the Academy of Sciences of the Republic of Uzbekistan Republic of Uzbekistan, Tashkent, Uzbekistan

+ rifqat@gmail.com

**Abstract.** This paper analyzes two approaches to the prediction of cardiovascular diseases, which have a binary classification problem based on the same Cleveland benchmark and aim to identify the advantages of quantum computing over traditional classifiers. This is a hybrid quantum random forest (HQNN, HQRF) for predicting the development of coronary heart disease in the early stages and an explainable prediction of cardiovascular diseases by the ensemble-quantum learning of Bagging-QSVC. The first approach uses various feature selection methods, which usually require a lot of computational effort, while the HQRF model and the earlier HQNN model are high-speed algorithms due to the nature of quantum computing. Numerical results show that HQRF is more suitable for small datasets, while HQNN is better suited for large datasets. In the second approach, the Bagging-QSVC model uses a quantum support vector classifier as a basic classifier. The results of the model are explained through the importance of the contribution of each individual feature using the Shaply (SHAP) algorithm. Comparative studies of other quantum classifiers on the Cleveland benchmark show the superiority of Bagging-QSVC with an accuracy of 90.16%. It follows from this that quantum machine learning classifiers are more effective than classical machine learning classifiers in predicting diseases of the cardiovascular system.

**Keywords:** machine learning, ensemble learning, quantum machine learning, explainable machine learning, heart disease prediction.