Основные этапы предлагаемой нами модели раннего прогнозирования и диагностики основных неблагоприятных сердечно-сосудистых событий показаны в Рис 1. Первым этапом предлагаемой модели является предварительная обработка данных набора данных KAMIR-NIH. На этапе предварительной обработки мы прошли через метод выбора признаков [25], в котором мы удалили неважные признаки из исходного набора данных и использовали наиболее важные признаки в качестве основного вклада в эту модель прогнозирования. Мы применили однократное горячее кодирование и кодирование меток [26] к выбранным признакам и подготовили наш предварительно обработанный набор данных для реализации модели. Предварительно обработанные данные разбиваются на обучающий набор данных (70 %) и проверочный набор данных (30 %) для обучения и тестирования модели соответственно. Вторым шагом предлагаемой нами модели является обучение модели прогнозирования на основе машинного обучения с использованием предварительно обработанного набора данных. На этом этапе обучения мы применили три различные модели машинного обучения в качестве моделей прогнозирования, например, случайный лес, дополнительное дерево и машину градиентного бустинга, и объединили их для создания ансамблевой модели для наилучшего прогнозирования и диагностики основных неблагоприятных сердечно-сосудистых событий. В предложенном нами классификаторе ансамбля мягкого голосования мы использовали алгоритмы машинного обучения случайного леса, дополнительного дерева и градиентного бустинга в качестве базовых классификаторов и скорректировали гиперпараметры с помощью алгоритма поиска по сетке для обучения этой модели, а затем оценили с помощью 5-кратной стратифицированной перекрестной проверки. Для обучения предложенной модели мы скорректировали веса этих классификаторов, т.к. этот классификатор голосования показал наилучшие результаты по удельному значению веса. Кроме того, мы использовали мягкое голосование для нашей модели. Мы настроили допуск, проверочную фракцию, вес и другие гиперпараметры в предложенной нами модели. Настройка гиперпараметров проиллюстрирована в Разделе 4.4. После обучения нашей ансамблевой модели на основе машинного обучения был применен тестовый набор данных (30%) для проверки производительности разработанной нами модели. После оценки модели на тестовых данных были извлечены лучшие значения гиперпараметров и доработана наилучшая модель прогнозирования путем корректировки гиперпараметров. Наконец, лучшие результаты модели прогнозирования будут извлечены и сравнены с результатами других моделей машинного обучения.

Для экспериментальной работы мы выбрали три алгоритма машинного обучения, названные случайным лесом [27, 28], дополнительным деревом [29] и машиной градиентного бустинга [30], и на основе этих трех базовых моделей разработали классификатор ансамбля мягкого голосования. По сравнению с другими алгоритмами машинного обучения, точность этих алгоритмов была сравнительно высокой, и это были лучше прогнозируемые модели для раннего прогнозирования и диагностики острого коронарного синдрома.

Разработанный нами классификатор ансамбля мягкого голосования представляет собой комбинацию нескольких классификаторов, в которых решения принимаются на основе отдельных решений, которые объединяются на основе значений вероятности, чтобы указать, что данные принадлежат к определенному классу. В ансамбле мягкого голосования прогнозы взвешиваются на основе важности классификатора и объединяются для получения суммы взвешенных вероятностей. Целевая метка с наибольшей суммой взвешенных вероятностей выбирается потому, что она имеет наибольшее значение для голосования (Рис 2). Пользовательские весовые коэффициенты также могут использоваться для вычисления средневзвешенного значения, чтобы придать большую важность и вовлеченность какой-либо конкретной модели обучения (базового классификатора). В отличие от жесткого голосования, мягкое голосование дает лучший результат и производительность, поскольку в нем используется усреднение вероятностей [31]. Классификатор ансамбля мягкого голосования скрывает слабость отдельных базовых классификаторов и превосходит общие результаты за счет агрегирования нескольких моделей прогнозирования. Основная цель ансамблевых методов – уменьшить систематическую ошибку и дисперсию.