

Метод ранней диагностики смертельных заболеваний, основанный на ИИ и Больших Данных, способный значительно продлить активную жизнь

Душкин Р. В., Лелекова В. А., Максимов В.С., Золман О., Тетерин О.О.

Аннотация. Старение — это часть человеческой жизни, часто сопровождающаяся тяжёлыми заболеваниями. В наше время люди и вовсе иногда не доживают до биологического старения организма из-за несвоевременно диагностированных болезней. В этой статье проводится обзор методик определения и анализа старения организма. К сожалению, у методов раннего обнаружения сопутствующих старению заболеваний ещё нет технического оснащения, позволяющего реализовать их в полной мере. В настоящей работе представлен новый разработанный научной группой программно-аппаратный комплекс мониторинга здоровья, который анализирует биопараметры человека с помощью алгоритмов искусственного интеллекта. Актуальность предложенной системы неоспорима в силу использованных алгоритмов искусственного интеллекта, с помощью которых можно быстро и точно проанализировать большое количество данных, касающихся старения человека. Статья будет интересна разработчикам искусственного интеллекта, биостатистикам и учёным, работающим над определением старения человеческого организма.

Ключевые слова: искусственный интеллект, биомаркеры, нейронные сети, старение, старение человека, мониторинг здоровья, 4P Medicine, машинное обучение, биологический возраст, определение тяжёлых заболеваний.

1. Введение

Старение естественно для каждого живого организма, при этом оно зачастую сопровождается заболеваниями, влияющими как на тело, так и на психику человека. С биологической точки зрения, старение — это результат влияния метаболических ошибок и внешних стрессовых факторов на индивидуальное развитие организма [Moskalev A., 2019]. Для разных людей одного возраста темпы старения могут существенно отличаться, как и темпы старения систем и органов в пределах одного организма — деградация одной системы вызывает изменения во многих других [Fedintsev A., 2017].

Над целью продления человеческой жизни и определения биохимических процессов, происходящих во время старения, трудится множество учёных. Для каждого организма старение происходит индивидуально, поэтому однозначного способа отследить сам процесс всё ещё не найдено. Тем не менее, существует множество разных способов определения биологического возраста, отличных от возраста календарного. Данные, идентифицирующие старение организма, называются биологическими маркерами старения или просто биомаркерами старения. Биомаркеры могут быть найдены во всём организме: в микрофлоре органов, в клетках, а, особенно, в цепочках ДНК и РНК. О биологическом возрасте и методах его определения, в том числе с помощью искусственного интеллекта (ИИ), машинного обучения и искусственных нейронных сетей, пойдёт речь во второй части настоящей статьи.

В связи с глобальным изменением образа жизни за последние десятилетия, преждевременные смерти регистрируются гораздо чаще. Согласно статистике, многие люди умирают значительно раньше физической смерти своих органов. Основной причиной являются тяжёлые заболевания, которые не были вовремя диагностированы и вылечены [Kupryushin A.S., 2016]. Обычно врачи проводят комплексное обследование только если стандартные анализы не соответствуют нормальным значениям, в то время как большинство болезней, протекающих латентно, не отслеживаются с помощью типовых тестов, доступных на сегодняшний день [Vasilkov V.G., 2017].

Применяющиеся повсеместно методы зачастую не помогают в обнаружении заболевания на ранних стадиях. Кроме того, из-за ограничений по стоимости всё меньше врачей пользуется современным оборудованием, таким как масс-спектрометрические приборы, для анализа биоматериалов. Учитывая всё вышесказанное, можно с уверенностью заявить — существует потребность в системах мониторинга здоровья, в которых данные могут быть проанализированы с помощью алгоритмов искусственного интеллекта. О разработанной совокупности решений повествует третья часть настоящей статьи.

2. Методы и данные

2.1. Биомаркеры и биологическое старение

Биомаркеры старения — это молекулярные, клеточные и физиологические параметры организма, которые прогнозируемо изменяются с возрастом — качественно или количественно [Moskalev A., 2019]. В научной среде такая единица измерения для биологического возраста считается наиболее актуальной и перспективной для ранней диагностики, прогнозирования и мониторинга хронических, а в том числе и возрастных, заболеваний. Многие патологии долгое время развиваются в латентной форме - обнаружить их на ранней стадии крайне сложно, поскольку их проявления неспецифичны для общего клинического случая. Биомаркеры дают возможность определить отклонение от нормы, связанное с риском развития конкретного заболевания, и назначить эффективную профилактику на ранней стадии [Pyrkov T.V., 2019].

Согласно Батлеру и др. [Butler R.N., 2004] биомаркер должен:

- изменяться с возрастом;
- предсказывать смерть лучше, чем календарный возраст;
- определять ранние стадии конкретной патологии, в частности — возрастного заболевания;
- быть минимально инвазивным — не требовать серьёзного хирургического вмешательства или болезненной процедуры.

Позже, Москалёв [Moskalev A., 2019] дополнил список:

- иметь высокую чувствительность к ранним признакам старения организма;
- быть предсказуемым в обозримом временном диапазоне;
- иметь низкую аналитическую вариативность — быть надёжным и воспроизводимым.

На сегодняшний день наиболее полной базой данных биомаркеров старения человека является Digital Aging Atlas [Craig T., 2015].

2.2. Практическое определение биомаркеров и биологического возраста

Проблематика биологического старения была представлена Вейтсманом и др. [Veytsman B., 2019] в качестве траектории в многомерном пространстве, каждое измерение которого представляет собой характеристики функционирования или работы организма. Эти траектории различны у разных людей и, более того, каждый человек следует по ним с определённой скоростью, которая может увеличиваться или уменьшаться в зависимости от конкретных обстоятельств. Фактически, каждая траектория может быть отображена в наборе естественных координат, в рассматриваемом случае — биомаркеров. Следовательно, появляется возможность сравнивать точки на конкретном наборе траекторий, то есть сравнивать отдельных людей и приходить к пониманию того, как строится траектория человека и почему изменяется скорость движения человека по собственной траектории.

Основной проблемой в определении такой «траектории» является двойственность подходов и потребность в серьёзном техническом оснащении. Различные подходы делятся на «теоретический» и «практический». Первый использует только подтверждённые данные, поэтому ограничен, поскольку текущие научные знания о человеческом организме пока ещё далеки от полноты. Второй учитывает все данные, но требует объёмных расчётов, из-за чего приходится прибегать к методам машинного обучения для анализа большого количества информации и отсеивания бесполезных выборок. Следовательно, для построения верных корреляций для выбранных биомаркеров необходимо соответствующее объёмам данных и сложности расчётов техническое

оснащение. Вариант программно-аппаратного комплекса, разработанного для анализа таких данных, представлен в третьей части этой статьи.

2.3. Перспективные биомаркеры

Хроматин. Основа хромосом, именно в составе хроматина происходит реализация генетической информации, репликация и репарация ДНК. Комплексное соединение ДНК, белков и РНК; наименьший компонент — нуклеосома. Существует множество модификаций, в частности — гетерохроматин, характеризующийся конденсированным (компактным) состоянием и низкой способностью к синтезу РНК.

Влияние на старение и заболевания:

- Доказано, что уменьшение гетерохроматина в организме ведёт к ряду хронических возрастных заболеваний (например — синдрому Вернера [Shumaker D.K., 2006; Zhang W., 2015]).
- Предположения о связи старения и уменьшении количества гетерохроматина в организме человека строятся на двух факторах. Первый из них — связь митохондриального стресса (результат неправильного конформационного сворачивания белков, приводящий к изменению работы различных генов), при котором уменьшается концентрация гетерохроматина в крови, и старения [Tian Y., 2016]. Второй — тенденция научных исследований [Scaffidi P., 2006; Larson K., 2012; Wood J.G., 2010].
- Ещё одно подтверждение связано с накоплением нуклеосом (продукта распада хроматина, его наименьшая составная часть) во время старения у млекопитающих [O'Sullivan R.J., 2010].

Модификации гистонов. Гистоны — белки, участвующие в упаковке нитей ДНК и регуляции таких процессов как образование РНК, создание дочерних ДНК и исправление повреждений в молекулах ДНК. Модификации гистонов — химические изменения в изначальных белках, которые влияют на образование РНК.

Влияние на заболевания и старение:

- Модификации гистонов, получаемые с помощью нуклеосом, накапливаются при старении [Contrepois K., 2017; Piazzesi A., 2016].
- Вариативность модификаций гистонов изменяется в процессе старения [Benayoun A., 2015].
- В качестве основных маркеров старения используются метилированные и ацетилированные гистоны. Количество метилированных гистонов снижается вместе с гетерохроматином при том же синдроме Вернера [Zhang W., 2015], в то время как количество ацетилированных коррелирует с возрастным ухудшением когнитивных функций [Peleg S., 2010].

Метилирование ДНК. Реакция метилирования происходит по CpG-динуклеотидам — участкам ДНК, играющим ключевую роль в активации врождённого иммунитета у позвоночных [Branda R.F., 1993]. По цепи ДНК CpG обычно распределены неравномерно, и, тем не менее, встречаются последовательности идущих друг за другом участков CpG — CpG-острова [Schubeler D., 2015]. Большинство отдельных CpG метилированы, за исключением CpG-островов [Neri F., 2017; Yang X., 2014].

Влияние на заболевания и старение:

- Метилирование CpG-островов является особенностью некоторых раковых клеток [Deaton A.M., 2011].
- По CpG-островам возможно отследить возрастное изменение у млекопитающих [Fraga M.F., 2007].
- Практические достижения в разработке «биологических часов» на основе реакции ДНК-метилирования. Некоторые группы учёных смогли измерить хронологический возраст человека [Horvath S., 2018; Nevalainen T., 2017; Wagner W., 2017], а группа Левина и др. [Levine M.E., 2018] смогла разработать часы, предсказывающие биологический возраст, используя алгоритмы множественной линейной регрессии и машинное обучение, способные анализировать и прогнозировать продолжительность жизни.
- Подтверждено, что данные, полученные при анализе метилированных CpG-групп, точны до уровня 0.95 корреляционного коэффициента между предсказанным возрастом и настоящим [Horvath S., 2018].

Микро-РНК. Класс некодирующих (не переводящихся в белок) РНК, вызывающих деградацию тканей [Bartel D.P., 2004].

Влияние на заболевания и старение:

- Следующими научными группами показано влияние на старение: мышечных клеток [Singh J., 2016; Drummond M.J., 2011], половых клеток [Battaglia R., 2016], крови [Olivieri F., 2017] и сыворотки крови [Smith-Vikos T., 2016] человека.
- Показано существование и доказано влияние двух микро-РНК (let-7 и mir-34) на повреждение ДНК [Kato M., 2009; Kato M., 2011], нейродеградацию [Lehmann S.M., 2012], старение [Tazawa H., 2007] и гибель клеток [He L., 2007].
- С помощью указанных выше микро-РНК есть возможность измерить продолжительность жизни [Ewald C.Y., 2016; Fitzenberger E., 2014].
- Показан процесс влияния микро-РНК на старение организма через центральный проводник сигналов старения [Tatar M., 2003] — сигнальный путь инсулина [Inukai S., 2012].
- Микро-РНК крайне перспективны как биомаркеры в силу: мобильности [Hoу A.M., 2012; Arroyo J.D., 2011; Hunter M.P., 2008; Vickers K.C., 2011], стабильности, возможности к количественному определению, защите от деградации [Huang X., 2013] и лёгкости в обнаружении [Machida T., 2015; Weber J.A., 2010].
- Самый точный метод выделения РНК неинвазивным путём — из сыворотки крови [Cheng L., 2014].

Микрофлора. Совокупность микроорганизмов, находящихся в симбиозе с человеком.

Влияние на заболевания и старение:

- Критически отличается у разных возрастных категорий, но меняется постепенно, со временем [Mangiola F., 2018], что является важным фактором для биомаркера.
- Неактуально для анализа данных пожилых людей из-за слишком большой вариативности [Claesson M.J., 2012].

Визуальное изображение.

Влияние на заболевания и старение:

- Используется в качестве количественной оценкой фенотипов и определения биомаркеров по базам данных [Zhao Q., 2016].
- Группами учёных [Gunn D.A., 2008] доказано, что воспринимаемый по изображению возраст человека не только считается биомаркером, но и связан с вариациями человеческих генов [Liu F., 2016].
- Биомаркером также считается постепенно проявляющаяся с возрастом пигментация лица, сильно связанная с некоторыми возрастными заболеваниями, например, с атеросклерозом сонных артерий [Miyawaki S., 2016].
- При помощи МРТ и снимков черепа, было доказано, что плотность тканей черепа [Colcombe S.J., 2003] и объём мозга [Driscoll I., 2009] снижаются при старении, что согласуется со снижением когнитивных функций человека с возрастом.

2.4. Сопоставление современных и клинических методов

Биологические исследования, включающие большое количество образцов, являются дорогостоящими и логистически сложными. Необходимая оптимизация процесса заключается в возможности получить достаточно полную базу данных для описания процесса старения не только качественно, но и количественно. Ещё одна проблема заключается в сильном отличии биологических и клинических маркеров оценки здоровья человека [Mitnitski A., 2019], однако и она решается.

Интеграция большого количества биомаркеров в клинические исследования возможна с помощью индекса хрупкости — наиболее простого и надёжного из существующих способов. Индекс хрупкости был представлен в 2001 году [Mitnitski A.B., 2001] как средство количественной оценки общего состояния здоровья у пожилых людей, а затем распространён и на взрослых людей [Rockwood K., 2011]. Индекс хрупкости определяется как соотношение числа нарушений здоровья человека, накопленных человеком, к общему числу нарушений в базе данных или исследовании [Mitnitski A., 2015]. Согласно более ранним оценкам, полученным с помощью поперечного демографического анализа, нарушения накапливаются экспоненциально, со скоростью в 3.5 % в год

[Mitnitski A., 2005]. Также, Минитски и др. было показано, что индекс хрупкости работает с клеточными биомаркерами воспаления, клеточного старения и генетическими маркерами [Zhavoronkov A., 2013].

Однако для такого простого способа возникло противоречие — индекс хрупкости учитывает любой недостаток здоровья, но не учитывает его характеристики, например, заболевания, инвалидность и симптомы. Это удобно для практического использования, однако противоречит первичным принципам клинического обучения, при котором в первую очередь важна точность диагностики заболевания, поскольку то, что помогает при состоянии А, может быть вредным при аналогичном состоянии Б. Самым простым примером подобного является диабет — введение инсулина поможет пациенту с излишним количеством сахара в крови, но будет вредно пациенту с недостатком сахара в крови.

Тем не менее, как количественная оценка старения индекс хрупкости полностью себя оправдывает, ведь не важно, какие конкретно накопились нарушения, важно, как они повлияли на организм. Так, для примера, как могут быть эквивалентны проблемы с кожей и сердечный приступ в количественном смысле? Не каждый сердечный приступ смертелен и не каждая сыпь может оказаться доброкачественной. В той степени, в которой такие нарушения влияют на организм, они и добавляются к количеству нарушений. Индекс хрупкости отражает фактический урон человеческому организму, не учитывая, каким именно заболеванием был нанесён ущерб, и, поэтому, как количественная характеристика, крайне практичен.

2.5. Методы искусственного интеллекта для изучения старения

Оценка старения — это первый шаг к принятию мер по снижению заболеваемости, социального и экономического бремени, связанного со старением [Belsky D.W., 2017]. Одной из основных гипотез является предположение о том, что хронические заболевания, ассоциирующиеся с возрастом, имеют общие компоненты генетической архитектуры, а следовательно — связаны со старением и оценкой рисков для здоровья в целом.

Тем не менее, для того чтобы медицина долголетия официально рассматривалась как отрасль медицины, её должны практиковать врачи.

Клиническая практика требует многого — клинических протоколов и руководств по диагностике и лечению с определёнными показателями результатов, официальных биомаркеров и лекарственных препаратов, одобренных регулируемыми органами власти. Для разработки хотя бы предварительных клинических рекомендаций необходимо отслеживать старение и рассматривать его как медицинское состояние, проводя специальные исследования для проверки эффективности и безопасности конкретных вмешательств в процесс.

В первую очередь клинические испытания предназначены для прогнозирования долгосрочных последствий применения фармацевтических или медицинских препаратов [Abbas I., 2016]. Сложность такого процесса заключается не столько в денежных затратах, сколько во временных. Пациенты умирают до окончания срока исследований, при этом сами исследователи также стареют [Loseva P., 2020]. Тем не менее, учёные работают над этой проблемой, используя искусственный интеллект, ведь, если строить общую модель исследования и проводить оставшиеся расчёты с её помощью, то клинические испытания займут недели, а не годы [Holford N.H.G., 2000]. Благодаря быстро растущему объёму медицинских данных, доступных исследователям, включая данные, получаемые из электронных медицинских карт и носимых устройств, сложные алгоритмы машинного обучения способны сэкономить миллиарды долларов, ускорить развитие медицины и расширить доступ к экспериментальным методам лечения [Woo M., 2019].

Наиболее точные методы расчёта биологического возраста являются предметом постоянных дебатов. Последние исследования показывают, что набор биомаркеров, а не какой-либо отдельный биомаркер, является наиболее эффективным средством оценки состояния здоровья пациента [Liu Z., 2018].

Для анализа такого большого количества данных больше всего подходят методы ИИ и машинного обучения. В настоящее время растёт число инструментов на основе ИИ, которые, имея доступ к соответствующим параметрам здоровья, используют различные аспекты

состояния здоровья и темпов старения пациента для составления прогноза [Zhavoronkov A., 2021].

Популярные модели биологического возраста обучены предсказывать хронологический возраст, однако часто не могут полностью отразить признаки возможного заболевания. Этот недостаток может быть устранён с помощью лог-линейных моделей риска, позволяющих применять линейную регрессию к задаче и использующих клинические данные [Mamoshina P., 2018].

Например, Мамошиной и др. [Mamoshina P., 2016] были использованы глубокие нейронные сети (DNN — Deep Neural Networks) для построения часов старения. Архитектура DNN перспективна в настоящее время в силу их способности определять скрытые закономерности в наборах данных и изучать многомерные данные в нетипичном представлении [Putin E., 2016]. Эта группа учёных рассматривала предсказание возраста как задачу регрессии, то есть полученная модель берёт вектор значений показателей из состава анализа крови и выдаёт одно значение возраста пациента. Чтобы решить проблему интерпретируемости DNN и получить более глубокое понимание данных, учёные применили анализ важности признаков (permutation feature importance или PFI) для ранжирования входных маркеров крови в соответствии с их важностью для прогнозирования возраста.

В другой работе [Bottou L., 2012] также рассматривали предсказание возраста как задачу регрессии и для оценки эффективности метода использовали стандартный коэффициент детерминации и точность ϵ -предсказания. Метод оценки одиночной ДНК показал себя перспективным для дальнейшего исследования в силу высокой точности.

В настоящей работе для обучения моделей машинного обучения (в том числе используемых в разработке авторского коллектива нейронных сетей) использовались описанные далее алгоритмы, хотя разработка системы для оценки показателей старения ими не ограничиваются:

- Стохастический оптимизатор градиентного спуска — метод оптимизации функции с подходящими свойствами. Фактически это оптимизация градиентного спуска, поскольку, в случае стохастического оптимизатора используется не весь набор данных, а случайно

выбранный набор подмножества данных, что сильно снижает вычислительную нагрузку, обеспечивая более быстрые итерации в обмен на более низкую скорость сходимости [Kingma D.P., 2015].

- Адаптивная оптимизация оценки момента — метод, вычисляющий адаптивные скорости обучения нейросети для разных параметров [Breiman L., 2001]. Прост в реализации, эффективен и подходит для задач с колеблющимися параметрами (в описываемом случае — биомаркерами).
- Оптимизация распространения среднеквадратичного значения.
- Линейная регрессия — простая оценка значений коэффициентов, используемых при анализе, с использованием имеющихся данных. Полезна для большого количества данных.
- Логистическая регрессия — оценка вероятности определённого значения с имеющимися параметрами. Полезна для большого количества данных.
- Методы Монте-Карло — подмножество вычислительных алгоритмов, которые используют процесс многократной случайной выборки для численных оценок неизвестных параметров. Они позволяют моделировать сложные ситуации, в которых задействовано много случайных величин, а поэтому и оценивать возможные риски.
- Марковские модели: марковская цепь, скрытая модель Маркова, марковский процесс принятия решений, частично наблюдаемый марковский процесс принятия решений.
- Трансформер — модель глубокого обучения, которая использует механизм внимания, дифференцированно взвешивая значимость каждой части входных данных. Распараллеливание обучения в модели трансформера даёт возможность проводить обучение на больших наборах данных.
- Метод опорных векторов (SVM) — линейный алгоритм, используемый в задачах классификации и регрессии. Задачей алгоритма является нахождение гиперплоскости, разделяющей полученные данные на два класса. Основным плюсом — возможность работы с большим объёмом данных.

- Линейный SVC — оптимизация ядра алгоритма SVM, позволяющая сократить количество расчётов.
- Алгоритм k-Nearest Neighbors — простой алгоритм, который классифицирует все данные или наблюдения на основе сходства друг с другом. Полезен тем, что легко поддается параллельной реализации и использует только локальную информацию, а следовательно — адаптивен.
- Алгоритм Naïve Bayes — упрощение классификатора Байеса, алгоритм, рассчитывающий вероятности, делая допущение об условной независимости величин друг от друга. Благодаря этому алгоритм «Наивного Байеса» работает крайне быстро.
- Простой перцептрон — алгоритм контролируемого обучения бинарных и мультиклассовых классификаторов. Полезен своей адаптивностью в работе с разделением классов.
- Классификатор дерева решений — алгоритм контролируемого машинного обучения, который используется для решения задач классификации. Фактически это разветвляющаяся цепочка вопросов «Да» или «Нет». Такой алгоритм не требует глобальной подготовки данных, а затраты памяти на реализацию вывода логарифмически зависима от количества данных, используемых для обучения, поэтому большое количество данных не сильно влияет на скорость вывода.
- Random Forests — алгоритм, заключающийся в использовании ансамбля деревьев решений [Hastie T., 2009]. Быстр, прост в реализации, адаптивен, устойчив к нелинейным функциям, способен обрабатывать неструктурированные данные.
- Gradient Boosting Classifier — алгоритм машинного обучения для задач классификации, который строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений [Ke H., 2017]. Интуитивно понятный и эффективный в прогнозировании алгоритм.
- Ridge Classifier — алгоритм классификации, позволяющий работать не только при линейных отличиях классов. С его помощью появляется

возможность избежать чрезмерной подгонки данных под значения классов.

- Bagging Classifier — алгоритм, оценивающий ансамбль решений и подгоняющий базовые классифицирующие значения (каждый на случайном подмножестве), а затем объединяющий их, а отдельные прогнозы используются для формирования окончательного вывода. Сильно уменьшает дисперсию выходных данных.

3. Результаты и обсуждение

Использование алгоритмов искусственного интеллекта для прогнозирования продолжительности и качества человеческой жизни находится в самом начале своего развития. Задачей настоящей работы было создание программно-аппаратного комплекса мониторинга здоровья, который сможет проанализировать биопараметры человека с помощью алгоритмов искусственного интеллекта.

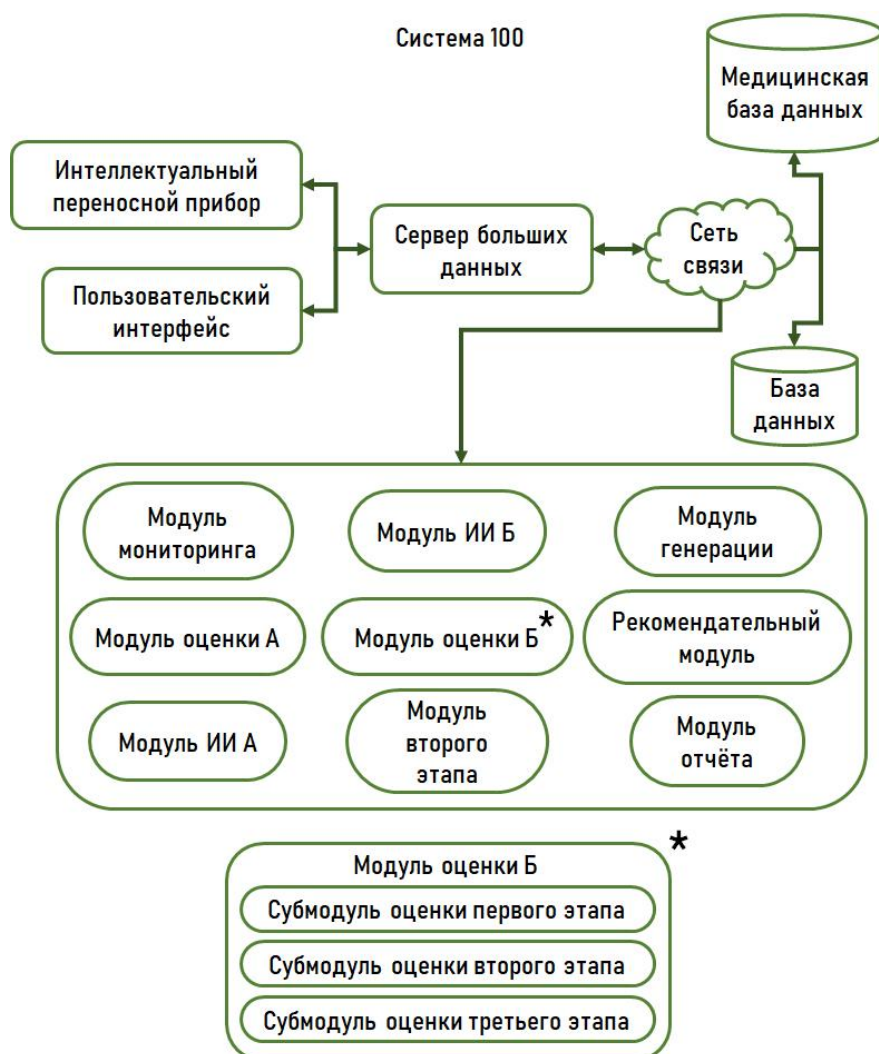


Рисунок 1. Блок-схема системы прогнозирования для оценки продолжительности и качества жизни

На рис. 1 показана блок-схема разрабатываемой Системы прогнозирования для оценки продолжительности и качества жизни, а в табл. 1 представлены описания конкретных модулей. Система прогнозирования сконфигурирована так, чтобы оценка жизни и общего состояния здоровья человека происходила на основе как можно большего количества факторов, влияющих на человека. Алгоритмы ИИ и, в частности, машинное обучение, вносят серьёзный вклад в:

- мониторинг системой множества параметров здоровья человека в режиме реального времени, включая динамическое изменение параметров;
- мониторинг состояния аппаратного обеспечения Системы прогнозирования;
- работу рекомендательного модуля, предлагающего шаги для улучшения состояния человека на основе аналогичных инцидентов из базы данных и предотвращения вероятных негативных сценариев ещё на доклинической стадии;
- корреляцию аномалий в базе данных с целью создания подкласса для уникальных ситуаций, их причин и последствий.

Тем не менее, для организации процесса сбора данных и обучения системы необходимо решить множество технических и организационных проблем. Среди технических — задача формирования обучающей выборки и целевого значения продолжительности жизни (фактора здоровья человека). Система прогнозирования также имеет потребность в формулировании критерия для оценки здоровья существующих наборов пациентов. Используемый в настоящей работе фактор здоровья является синтетической мерой, зависящей от количества системных заболеваний, текущих значений тестов и функциональных характеристик организма.

Входные данные Системы прогнозирования представляют собой набор параметров, полученных из медицинских карт, переносных устройств, анкет и других источников, включая загрузку лабораторных

анализов биоматериалов на сервер самой Системы. Выходные данные Системы прогнозирования представлены в виде отчёта о состоянии организма, возможных заболеваний и «фактора здоровья» человека, который напрямую связан с продолжительностью жизни. Чем выше этот фактор, тем дольше и качественнее прогнозируется жизнь человека.

Таблица 1. Описание модулей запатентованного метода системы прогнозирования

Название модуля	Описание модуля
Интеллектуальный переносной прибор	Содержит по меньшей мере один датчик для регистрации биопараметров, например Garmin (возможна реализация на платформе агрегации, такой как HealthKit, GoogleFit или другие мобильные приложения для регистрации биопараметров).
Пользовательский интерфейс	Позволяет пользователю просматривать данные различными методами, такими как графики, диаграммы и таблицы.
Сервер больших данных	Данные собраны из медицинской базы данных и одобрены Всемирной организацией здравоохранения и (или) Всемирной медицинской ассоциацией и (или) другими подобными ассоциациями. Список параметров на сервере больших данных, связанных с конкретными заболеваниями, может меняться в зависимости от новых решений Всемирной организации здравоохранения и других ассоциаций.
Сеть связи	Связь с сервером больших данных.
База данных	Хранит множество параметров здоровья большого количества людей.

Медицинская база данных	При совмещении с обычной базой данных может являться реляционной базой данных, распознающей взаимосвязи между имеющимися в обоих базах данных элементами.
Модуль мониторинга	Настроен на учёт и перераспределение множества параметров здоровья.
Модуль оценки А	Модель искусственного интеллекта, обученная на данных, полученных из базы данных.
Модуль искусственного интеллекта А	Структурирует персональные наборы данных множества индивидуумов и формирует цифровой профиль на основе совпадений, полученных из, по меньшей мере, одного набора данных пользователя и множества наборов данных общей популяции, полученных из баз данных.
Модуль искусственного интеллекта Б	<ol style="list-style-type: none"> 1. генерирует список необходимых параметров для постоянного отслеживания возможных заболеваний; 2. разрабатывает персонализированные алгоритмы для каждого отчёта о состоянии здоровья пользователя; 3. при имеющейся информации о конкретном пользователе за продолжительный отрезок времени, использует собранные данные для обучения моделей ИИ с целью прогнозирования риска заболевания конкретного пользователя и создания персональных профилактических расчётов без использования общих наборов данных населения.

Модуль оценки Б	Оценивает обучающую выборку данных, чтобы сделать выводы на основе большого набора данных и определить по крайней мере одну характеристику из обучающей выборки.
Субмодуль оценки первого этапа	Мгновенная оценка данных о человеке, полученных в определённый момент времени.
Субмодуль оценки второго этапа	Модель машинного обучения для обработки исторических данных, которая анализирует историю заболеваний человека и обучается на ней.
Субмодуль оценки третьего этапа	<p>Задачами этого модуля являются:</p> <ol style="list-style-type: none"> 1. Распознавание большого количества сходных параметров в множестве входных параметров для оценки индивидуальной взаимосвязи между качеством жизни человека и его соответствующим качеством здоровья. 2. Формирование основной персонализированной модели машинного обучения, обученной на большой выборке данных и настроенной на конкретного человека в течение длительного периода времени.
Модуль второго этапа	Разрабатывает модель машинного обучения. Такая обученная модель анализирует все известные медицинские данные конкретного человека.
Модуль генерации	Предоставляет выходные данные, представляющих собой фактор оценки здоровья человека, непосредственно связанный с продолжительностью жизни.

Рекомендательный модуль	Предоставляет персональные рекомендации по профилактике тяжёлых заболеваний на разных этапах их развития.
Модуль отчёта	Создание отчёта о риске возможного заболевания.

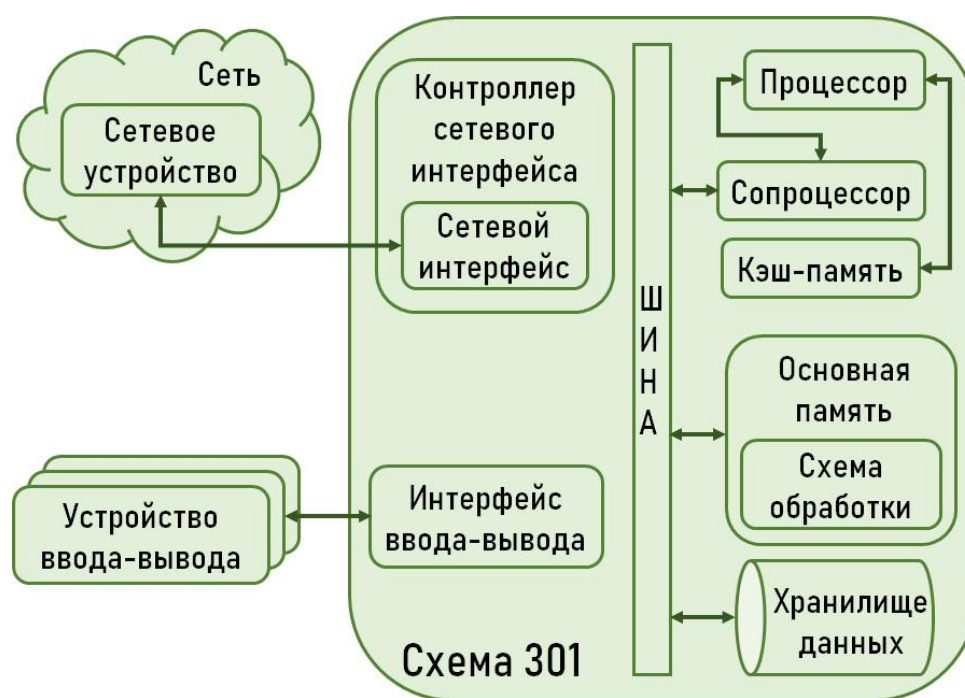


Рисунок 2. Блок-схема аппаратного обеспечения Системы прогнозирования

На рис. 2 представлена блок-схема аппаратного обеспечения Системы прогнозирования, включающая в себя:

- **процессор** для выполнения действий в соответствии с инструкциями, для загрузки инструкций работы из основной памяти (или хранилища данных) в кэш-память, для загрузки инструкций из кэш-памяти в бортовые регистры;
- **блок кэш-памяти** для хранения инструкций;
- **сопроцессор** для поддержки взаимодействия процессора и основной памяти;
- **шина** — интерфейс, обеспечивающий обмен данными между внутренними компонентами Системы прогнозирования;

- **блок основной памяти**, содержащий исполняемый компьютером код и включающий в себя схему обработки, которая сконфигурирована на основе ранее описанной схемы на рис. 1;
- **контроллер сетевого интерфейса** для управления одним или несколькими сетевыми интерфейсами и для подключения к сетевым устройствам (например, для доступа к сети);
- **интерфейс ввода-вывода** для облегчения отправки и приёма данных на различные устройства ввода-вывода (мобильные телефоны, принтеры и т. п.);
- **хранилище данных.**

Прогнозирование и оценка качества и продолжительности жизни человека происходят по шагам, показанным на блок-схеме, представленной на рис. 3, в соответствии с описанным ранее аппаратно-программным обеспечением.



Рисунок 3. Блок-схема способа прогнозирования качества и продолжительности жизни в настоящем изобретении

Первая база данных включает в себя различные биоданные, хранящиеся и собирающиеся в динамике индивидуально для каждого человека. В базе данных, при завершении первого шага находятся:

- данные общей популяции, на которых базируются научные статьи;

- собранные представленной системой данные, которые распределяются по схожести параметров (например, только мужчины, 45 лет, проживающие в мегаполисе, перенёсшие инсульт в прошлом, курящие и т. д.)
- данные, накопленные за 3-5 лет по каждому пользователю. Предполагается, что именно такой вариант позволит сделать наиболее точный прогноз траектории жизни человека.

На втором шаге биопараметры здоровья отслеживаются и оцениваются моделью машинного обучения, обученной на базе данных и медицинской базе данных. После того, как биопараметры преобразованы в набор персональных данных, модулем ИИ формируются цифровые профили пользователей, считающиеся обучающей выборкой для модели машинного обучения.

На третьем шаге оценивается обучающая выборка, чтобы сделать на её основе выводы и определить, по крайней мере, одну характеристику для описания взаимосвязи между качеством и продолжительностью жизни человека. При наличии медицинской истории человека, модель обучается на большем количестве данных, а, следовательно, в процессе работы распознаёт больше закономерностей между качеством жизни и уровнем здоровья индивида.

Затем на четвёртом шаге формируется обученная модель машинного обучения, которая анализирует множество биопараметров человека за разные периоды жизни, влияющие как положительно, так и отрицательно на прогноз о продолжительности жизни.

На последнем шаге происходит генерация выходных данных. В частности, коэффициент здоровья человека, который напрямую связан с продолжительностью жизни. Для примера — коэффициент здоровья говорит о большей продолжительности жизни, когда риск тяжёлых заболеваний оценён и невысок. Персональные рекомендации по профилактике тяжёлых заболеваний предоставляются лично человеку посредством сгенерированного отчёта при помощи модуля ИИ, сконфигурированного для извлечения необходимых данных из неструктурированного множества.

4. Заключение

Описанный в этой статье метод представляет собой систему 4P Medicine — Prevention, Prediction, Participatory, Personalized. Система отслеживает двадцать самых тяжёлых заболеваний на ранних известных стадиях и способна увеличить продолжительность и качество человеческой жизни. Ранний анализ биопараметров индивида, пока симптомы даже не видны по традиционным анализам (биохимия, УЗИ, МРТ и т.д.), но развитие болезни уже протекает, позволяет существенно сэкономить на дорогостоящем лечении, когда тяжёлая болезнь уже запущена. Более того, описываемая разработка оснащена предложенным программно-аппаратным комплексом и отслеживает более пятидесяти параметров с устройств носимой электроники в режиме реального времени, а также динамику состояния человека по более чем тремстам пятидесяти параметрам крови и других биоматериалов при регулярном измерении (свыше двух в год), предоставляя пользователю своевременные рекомендательные отчёты о состоянии здоровья. Как итог, описываемая разработка является большим подспорьем в поиске «эффекта бабочки» жизненной траектории каждого человека, когда всё ещё возможно внести необходимые изменения в жизнь людей, предоставляя им возможность прожить больше активных и счастливых лет без ограничений в виде тяжёлых заболеваний.

Литература

Abbas I. (2016) Modeling and simulation in clinical trials. Proceedings of the Modeling and Simulation in Medicine Symposium (MSM '16). Society for Computer Simulation International, San Diego, CA, USA, 1:1–6.

Arroyo J.D. et al. (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc. Natl. Acad. Sci. USA 108(12):5003–5008.

Bartel D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 116(2):281–297.

Battaglia R. et al. (2016) MicroRNAs are stored in human MII oocyte and their expression profile changes in reproductive aging. *Biol. Reprod.* 95(6):131.

Belsky D.W., Moffitt T.E., Cohen A.A., et al. (2017) Eleven telomere, epigenetic clock, and biomarker-composite quantifications of biological aging: Do they measure the same thing? *Am. J. Epidemiol.* 2017;kwx346 .

Benayoun B.A., Pollina E.A., Brunet A. (2015) Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nat. Rev. Mol. Cell. Biol.* 16(10):593–610.

Bottou L., Bousquet O. (2012) The Tradeoffs of Large Scale Learning. Sra, Suvrit; Nowozin, Sebastian; Wright, Stephen J. (eds.). *Optimization for Machine Learning*. Cambridge: MIT Press. pp. 351–368.

Branda R.F., Moore A.L., Mathews L., McCormack J.J., Zon G. (1993) Immune stimulation by an antisense oligomer complementary to the rev gene of HIV-1. *Biochem. Pharmacol.* 45(10):2037-2043.

Breiman L. (2001) Random Forests. *Machine Learning* 45:5–32.

Butler R.N. et al. (2004) Biomarkers of aging: from primitive organisms to humans. *J. Gerontol. A Biol. Sci. Med. Sci.* 59:B560.

Cheng L. et al. (2014) Exosomes provide a protective and enriched source of miRNA for biomarker profiling compared to intracellular and cell-free blood. *J. Extracell. Vesicles* 26:3.

Claesson M.J. et al. (2012) Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488(7410):178–184.

Colcombe S.J. et al. (2003) Aerobic fitness reduces brain tissue loss in aging humans. *J Gerontol. A Biol. Sci. Med. Sci.* 58(2):176–180.

Contrepois K. et al. (2017) Histone variant H2A.J accumulates in senescent cells and promotes inflammatory gene expression. *Nat. Commun.* 8:14995.

Craig T. et al. (2015) The Digital Ageing Atlas: integrating the diversity of age-related changes into a unified resource. *Nucleic. Acids. Res.* 43:D873.

Deaton A.M., Bird A. (2011) CpG islands and the regulation of transcription. *Genes Dev.* 25(10):1010–1022.

Driscoll I. et al. (2009) Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology* 72(22):1906–1913.

Drummond M.J. et al. (2011) Aging and microRNA expression in human skeletal muscle: a microarray and bioinformatics analysis. *Physiol. Genomics.* 43(10):595–603.

Ewald C.Y., Marfil V., Li C. (2016) Alzheimer-related protein APL-1 modulates lifespan through heterochronic gene regulation in *Caenorhabditis elegans*. *Aging Cell* 15(6):1051–1062.

Fedintsev A. et al. (2017) Markers of arterial health could serve as accurate non-invasive predictors of human biological and chronological age. *Aging* 9:1280.

Fitzenberger E. et al. (2014) The polyphenol quercetin protects the mev-1 mutant of *Caenorhabditis elegans* from glucose-induced reduction of survival under heat-stress depending on SIR-2.1, DAF-12, and proteasomal activity. *Mol. Nutr. Food Res.* 58(5):984–994.

Fraga M.F., Esteller M. (2007) Epigenetics and aging: the targets and the marks. *Trends. Genet.* 23(8):413–418.

Gunn D.A. et al. (2008) Perceived age as a biomarker of ageing: a clinical methodology. *Biogerontology* 9(5):357–364.

Hastie T., Tibshirani R., Friedman J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition. Springer Series in Statistics.

He L. et al. (2007) A microRNA component of the p53 tumour suppressor network. *Nature* 447(7148):1130–1134.

Holford N. H. G., Kimko H.C., Monteleone J.P.R., Peck C.C. (2000) Simulation of Clinical Trials. *Annual Review of Pharmacology and Toxicology* 40(1):209-234.

Horvath S., Raj K. (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19(6):371–384.

Hoy A.M., Buck A.H. (2012) Extracellular small RNAs: what, where, why? *Biochem. Soc. Trans.* 40(4):886–890.

Huang X. et al. (2013) Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genom.* 14:319.

Hunter M.P. et al. (2008) Detection of microRNA expression in human peripheral blood microvesicles. *PLoS ONE* 3(11):3694.

Inukai S. et al. (2012) Novel microRNAs differentially expressed during aging in the mouse brain. *PLoS ONE* 7(7):40028.

Kato M. et al. (2009) The mir-34 microRNA is required for the DNA damage response in vivo in *C. elegans* and in vitro in human breast cancer cells. *Oncogene* 28(25):2419–2424.

Kato M. et al. (2011) Age-associated changes in expression of small, noncoding RNAs, including microRNAs in *C. elegans*. *RNA* 17(10):1804–1820.

Ke G., Meng Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *NIPS*.

Kingma D.P., Ba, J. (2015). Adam: A Method for Stochastic Optimization. *CoRR*, abs/1412.6980.

Kupryushin A.S., Markova A.A., Kupryushina N.V., Vishnyakova Z.S., Latynova I.V., Semina M.N. СТРУКТУРА ПРИЧИН СМЕРТИ БОЛЬНЫХ ПОЖИЛОГО И СТАРЧЕСКОГО ВОЗРАСТА, УМЕРШИХ ДОМА // Современные проблемы науки и образования. – 2016. – № 3.

Larson K. et al. (2012) Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. *PLoS Genet* 8(1):e1002473.

Lehmann S.M. et al. (2012) An unconventional role for miRNA: let-7 activates Toll-like receptor 7 and causes neurodegeneration. *Nat. Neurosci.* 15(6):827–835.

Levine M.E. et al. (2018) An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* 10(4):573–591.

Liu F. et al. (2016) The MC1R gene and youthful looks. *Curr. Biol.* 26(9):1213–1220.

Liu Z., Kuo P.-L., Horvath S., Crimmins E., Ferrucci L., Levine M. (2018). Phenotypic age: a novel signature of mortality and morbidity risk. bioRxiv363291 [Preprint].

Loseva P. (2020) Против часовой стрелки / «Альпина Диджитал», 2020 — (Primus).

Machida T. et al. (2015) MicroRNAs in salivary exosome as potential biomarkers of aging. *Int. J. Mol. Sci.* 16(9):21294–21309.

Mamoshina P., Kochetov K., Putin E., Cortese F., Aliper A., Lee W.S., Ahn S.M., Uhn L., Skjodt N., Kovalchuk O., Scheibye-Knudsen M., Zhavoronkov A. (2018) Population Specific Biomarkers of Human Aging: A Big Data Study Using South Korean, Canadian, and Eastern European Patient Populations. *J. Gerontol. A Biol. Sci. Med. Sci.* 73(11):1482-1490.

Mamoshina P., Vieira A., Putin E., Zhavoronkov A. (2016) Applications of deep learning in biomedicine. *Mol. Pharm.* 13:1445–1454.

Mangiola F. et al (2018) Gut microbiota and aging. *Eur. Rev. Med. Pharmacol. Sci.* 22(21):7404–7413.

Mitnitski A., Rockwood K. (2015) Aging as a process of deficit accumulation: its utility and origin. *Interdiscip. Top. Gerontol.* 40:85–98.

Mitnitski A., Song X., Skoog I., Broe G.A., Cox J.L., Grunfeld E., Rockwood K. (2005) Relative fitness and frailty of elderly men and women in developed countries and their relationship with mortality. *J. Am. Geriatr. Soc.* 53(12):2184–2189.

Mitnitski A., Rockwood K. (2019) The Problem of Integrating of Biological and Clinical Markers of Aging. In: Moskalev A. (eds) Biomarkers of Human Aging. Healthy Ageing and Longevity, vol 10. Springer, Cham.

Mitnitski A.B., Mogilner A.J., Rockwood K. (2001) Accumulation of deficits as a proxy measure of aging. *Sci. World J.* 1:323–336.

Miyawaki S. et al. (2016) Facial pigmentation as a biomarker of carotid atherosclerosis in middleaged to elderly healthy Japanese subjects. *Skin. Res. Technol.* 22(1):20–24.

Moskalev A., editor. Biomarkers of Human Aging. Springer International Publishing; 2019.

Moskalev A. (2019) Introduction. In: Moskalev A. (eds) Biomarkers of Human Aging. Healthy Ageing and Longevity, vol 10. Springer, Cham.

Neri F. et al. (2017) Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543(7643):72–77.

Nevalainen T. et al. (2017) Obesity accelerates epigenetic aging in middle-aged but not in elderly individuals. *Clin. Epigenetics* 9:20.

O’Sullivan R.J. et al. (2010) Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nat. Struct. Mol. Biol.* 17(10):1218–1225.

Olivieri F. et al. (2017) Circulating miRNAs and miRNA shuttles as biomarkers: perspective trajectories of healthy and unhealthy aging. *Mech. Ageing Dev.* 165(Pt B):162–170.

Peleg S. et al. (2010) Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science* 328(5979):753–756.

Piazzesi A. et al. (2016) Replication-independent histone variant H3.3 controls animal lifespan through the regulation of pro-longevity transcriptional programs. *Cell. Rep.* 17(4):987–996.

Putin E., Mamoshina P., Aliper A., Korzinkin M., Moskalev A., Kolosov A., Ostrovskiy A., Cantor C., Vijg J., Zhavoronkov A. (2016) Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging (Albany NY)*. 8(5):1021-33.

Pyrkov T.V., Fedichev P.O. (2019) Biological Age is a Universal Marker of Aging, Stress, and Frailty. In: Moskalev A. (eds) Biomarkers of Human Aging. Healthy Ageing and Longevity, vol 10. Springer, Cham.

Rockwood K., Song X., Mitnitski A. (2011) Changes in relative fitness and frailty across the adult lifespan: evidence from the Canadian National Population Health Survey. *CMAJ* 183(8):E487–E494.

Scaffidi P., Misteli T. (2006) Lamin A-dependent nuclear defects in human aging. *Science* 312(5776):1059–1063.

Schubeler D. (2015) Function and information content of DNA methylation. *Nature* 517(7534):321–326.

Shumaker D.K. et al. (2006) Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *Proc. Natl. Acad. Sci. USA* 103(23):8703–8708.

Singh J. et al. (2016) Aging-associated changes in microRNA expression profile of internal anal sphincter smooth muscle: role of microRNA-133a. *Am. J. Physiol. Gastrointest. Liver. Physiol.* 311(5):G964–G973.

Smith-Vikos T. et al. (2016) A serum miRNA profile of human longevity: findings from the Baltimore longitudinal study of aging (BLSA). *Aging (Albany NY)* 8(11):2971–2987.

Tatar M., Bartke A., Antebi A. (2003) The endocrine regulation of aging by insulin-like signals. *Science* 299(5611):1346–1351.

Tazawa H. et al. (2007) Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc. Natl. Acad. Sci. USA*. 104(39):15472–15477.

Tian Y. et al. (2016) Mitochondrial stress induces chromatin reorganization to promote longevity and UPR (mt). *Cell* 165(5):1197–1208.

Vasilkov V.G., Safronov A.I. (2017) МЕДИЦИНА КРИТИЧЕСКИХ СОСТОЯНИЙ: ИНТЕНСИВНАЯ ДИАГНОСТИКА. АКТУАЛЬНЫЕ ВОПРОСЫ ДИАГНОСТИКИ, ЛЕЧЕНИЯ И РЕАБИЛИТАЦИИ БОЛЬНЫХ Материалы XIX юбилейной межрегиональной научно-практической конференции, посвященной 40-летию ПИУВ - филиала ФГБОУ ДПО РМАНПО Минздрава России.

Veytsman B., Cui T., Baranova A. (2019) Practical Detection of Biological Age: Why It Is not a Trivial Task. In: Moskalev A. (eds) *Biomarkers of Human Aging. Healthy Ageing and Longevity*, vol 10. Springer, Cham.

- Vickers K.C., et al. (2011) MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell. Biol.* 13(4):423–433.
- Wagner W. (2017) Epigenetic aging clocks in mice and men. *Genome. Biol.* 18(1):107.
- Weber J.A., et al. (2010) The microRNA spectrum in 12 body fluids. *Clin. Chem.* 56(11):1733–1741.
- Woo M. (2019) An AI boost for clinical trials. *Nature* 573:100-102.
- Wood J.G., et al. (2010) Chromatin remodeling in the aging genome of *Drosophila*. *Aging Cell* 9(6):971–978.
- Yang X., et al. (2014) Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 26(4):577–590.
- Zhang W., et al. (2015) Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science* 348(6239):1160–1163.
- Zhao Q., et al. (2016) Dissecting the precise role of H3K9 methylation in crosstalk with DNA maintenance methylation in mammals. *Nat. Commun.* 7:12464.