

МНОГОМАСШТАБНОЕ ПОПУЛЯЦИОННОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ РАЗВИТИЯ И ЛЕЧЕНИЯ АРТЕРИАЛЬНОЙ ГИПЕРТЕНЗИИ

А. А. СЕМАКОВА¹, Н. Э. ЗВАРТАУ^{1,2}, С. В. КОВАЛЬЧУК¹, А. В. БУХАНОВСКИЙ¹

¹Университет ИТМО, 197101, Санкт-Петербург, Россия
E-mail: a.a.semakova@gmail.com

²Национальный медицинский исследовательский центр имени В. А. Алмазова, 197341, Санкт-Петербург, Россия

Проявление и развитие артериальной гипертензии (АГ) сопряжено с множеством эндогенных и экзогенных факторов. Представлена динамическая популяционная модель развития АГ в различных временных масштабах с учетом ориентированных на пациента стратегий лечения. Популяционная динамика группы пациентов с АГ построена с использованием демографической модели. На основе данных электронных медицинских карт выделены однородные группы пациентов с целью моделирования развития АГ на индивидуальном уровне. Для пациентов из каждой группы разработаны модели изменчивости количественных характеристик цифрового профиля пациента и оценены вероятности развития сопутствующих патологий. Построена модель тактики и стратегии антигипертензивного лечения на основе классификатора, прогнозирующего для пациента наиболее эффективный вид терапии по имеющимся до начала лечения индивидуальным компонентам его цифрового профиля. Разработанная технология моделирования может быть использована для оптимизации процессов системы здравоохранения, в первую очередь в части оказания амбулаторной помощи, для оценки эффективности внедрения новых препаратов и стратегий лечения, а также для моделирования сценариев введения новых клинических рекомендаций.

Ключевые слова: артериальная гипертензия, динамическая популяционная модель, цифровой профиль пациента, кластеризация, CART-алгоритм

Артериальная гипертензия (АГ) широко распространена: по данным ВОЗ, у 40 % населения РФ. Ее проявление и развитие сопряжено с множеством эндогенных и экзогенных факторов, вследствие чего существующие клинические рекомендации предусматривают большое количество вариантов лечения без четких различий в эффективности терапии. АГ требует пожизненной медикаментозной терапии, при этом эффективность лечения зависит от общего состояния пациента, сопутствующих заболеваний и их терапии. Существуют исследования, ориентированные на разработку клинических инструментов для оценки рисков появления и развития АГ на основе методов машинного обучения. В работе [1] на основе анамнеза пациента с помощью нейронной сети прогнозируется риск возникновения АГ. В статье [2] модель идентификации рисков развития АГ строится с использованием алгоритма случайных лесов (*random forest*). В исследовании [3] сравниваются 15 регрессионных предсказательных моделей риска развития АГ, в качестве предикторов выступают полученные с помощью доказательной медицины факторы риска. В работах [4, 5], риски развития АГ оцениваются с учетом причинно-следственных связей между АГ и сопутствующими заболеваниями. В целом перечисленные исследования ориентированы на выявление *самой возможности развития* АГ у конкретного, изначально здорового, человека. Однако в них не рассматривается ни многолетняя динамика развития самого заболевания, ни тактика и стратегия лечения конкретного пациента, ни популяционная изменчивость АГ (что принципиально важно для организации самой системы здравоохранения). Как следствие, целью настоящей статьи является изложе-

ние принципов построения и использования динамической популяционной модели развития АГ в различных временных масштабах с учетом ориентированных на пациента стратегий лечения.

В отличие от многих других нозологий АГ диагностируется по одному ключевому признаку — систематическому повышению артериального давления (АД) [6]. На основе уровня АД в клинической практике больные АГ классифицируются по степени тяжести нозологии (АГ 1-й, 2-й и 3-й степени), а также уровню риска сердечно-сосудистых осложнений [6]. В целом популяция больных АГ неоднородна; один и тот же больной может в разное время соответствовать разным классам в соответствии с динамикой заболевания. Исходя из этого для описания развития АГ на индивидуальном уровне можно ввести цифровой профиль пациента, детализированный в двух временных масштабах: годовом (макроуровень) и внутригодовом (микроуровень, с детализацией „минуты-часы“). На рис. 1 приведена феноменологическая модель развития АГ.

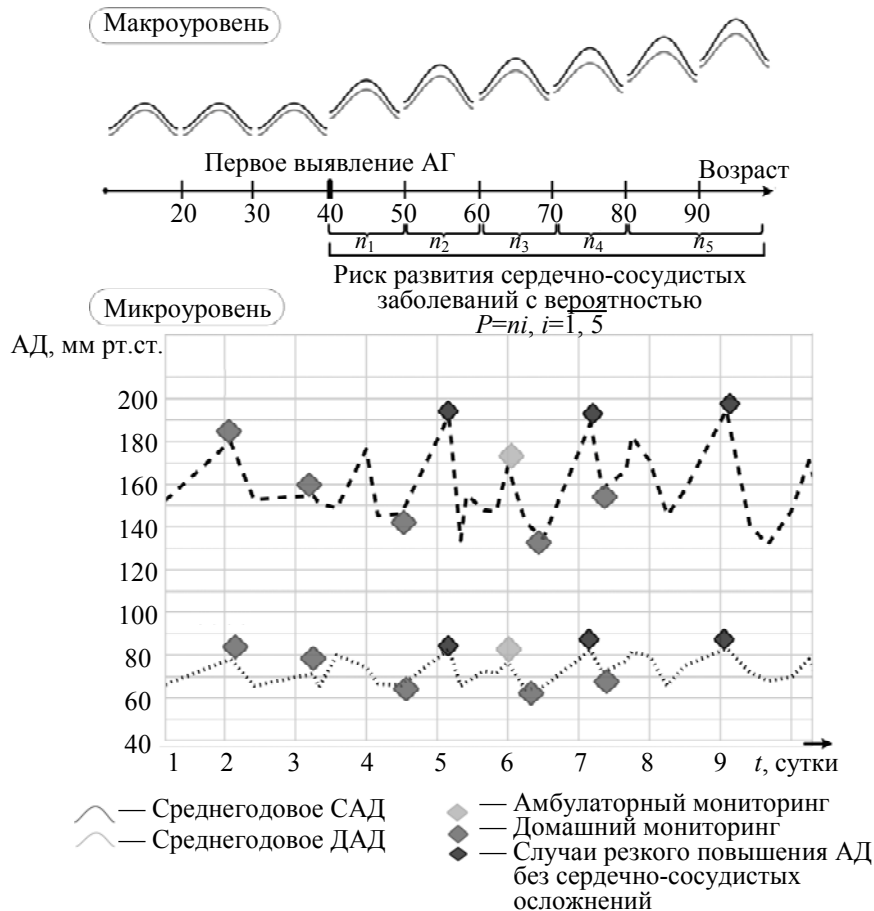


Рис. 1

Прогрессирование АГ и появление сопутствующих заболеваний могут быть описаны на макроуровне (который характеризуется среднегодовым значением АД, а также числом случаев превышения пороговых значений, включая экстремальные выбросы, например, гипертонические кризы). На микроуровне описываются внутрисуточные и межсуточные вариации АД, связанные с режимом сна и бодрствования, а также скачки АД патологического и непатологического характера (например, кратковременные превышения пороговых значений при нагрузках и стрессах). При этом количественные характеристики АД в цифровом профиле могут быть известны фрагментарно, в нерегулярных временных отсчетах. На практике они получают посредством измерений на амбулаторном приеме, суточного амбулаторного и домашнего мониторинга (дневники пациентов). Эти источники метрологически разнородны, однако совокупность их данных достаточна для статистического обобщения в рамках

популяционной модели, поскольку различия между пациентами более значимы, чем сама ошибка измерений.

С помощью цифрового профиля каждый пациент в произвольный момент времени может быть представлен кортежем из следующих групп переменных:

- наборы измерений систолического (САД) и диастолического (ДАД) АД по всем трем способам измерений (за всю историю наблюдений пациента);
- анамнез пациента (дата рождения, пол, национальность, наследственный базис, перенесенные заболевания, наличие/отсутствие вредных привычек и пр.);
- параметры, обладающие макроизменчивостью (например, индекс массы тела, ИМТ);
- текущий набор диагнозов;
- текущий набор медицинских назначений (процедуры и препараты).

Из-за отсутствия значимых причинно-следственных связей, влияющих на появление и эволюцию АГ в целом, популяционная модель развития и лечения строится непосредственно по данным, с использованием методов статистического анализа и машинного обучения. В качестве источников данных, достаточных для построения цифровых профилей, охватывающих все перечисленные переменные, может быть использована информация крупных медицинских центров, ведущих, в том числе, амбулаторный прием и консультирование. В настоящей работе использована база данных медицинской информационной системы ФГБУ „Национального медицинского исследовательского центра им. В.А. Алмазова“ Минздрава России, содержащая 4521 электронную медицинскую карту (ЭМК) пациентов, обратившихся с 2010 по 2015 г. за консультацией по причине повышения уровня АД. На основе технологии автоматического разбора медицинских текстов [7] из ЭМК пациентов были извлечены данные о факторах риска и поражениях органов-мишеней, которые были оценены у пациента на амбулаторном приеме. На основе извлеченных данных построены цифровые профили пациентов, которые затем использовались для работы с моделью.

На макроуровне общая динамика группы пациентов с АГ воспроизводится на основе демографической модели. Так, численность всей популяции заданного возраста в произвольный момент времени описывается дифференциальным уравнением [8] с учетом значений прироста и смертности населения каждый год. В нашем случае эти данные взяты из статистической отчетности по Санкт-Петербургу и Ленинградской области. Доля больных АГ в общей популяции выражается через отношение темпов приростов численности пациентов к численности здоровой популяции. При этом сам прирост численности больных АГ связан с возрастом [9]: установлено, что по медиане (т.е. как минимум, у половины пациентов) диагноз АГ ставится с 60 лет, однако 95 %-ный интервал наступления данного события — от 18 до 77 лет. Фиксация факта АГ зависит от гендерной принадлежности пациента: у мужчин АГ может появиться в более раннем возрасте (медиана — 56 лет), чем у женщин (медиана — 62 года). Важно отметить, что по этому массиву данных представленные параметры оценены с ошибкой в опасную сторону (так как обычно пациенты обращаются за консультацией в крупные центры, в течение некоторого времени наблюдаясь по месту жительства). Однако для модели это не принципиально, поскольку факт обнаружения АГ в ней рассматривается не как начало патологических изменений в организме, а как отсчет для последующего контроля динамики.

Для моделирования развития АГ у конкретного пациента необходимо учесть, что она может быть вызвана принципиально различными причинами. Как следствие, для построения модели требуется выделить однородные группы пациентов. Для этого был использован алгоритм *t-SNE* (*t-distributed stochastic neighbor embedding*) [10] применительно к пяти группам переменных цифрового профиля, описанных выше. В результате кластеризации в популяции пациентов с АГ выделено восемь групп с вероятностями: $P_1 = 0,07$; $P_2 = 0,08$; $P_3 = 0,12$; $P_4 = 0,09$; $P_5 = 0,22$; $P_6 = 0,2$; $P_7 = 0,15$; $P_8 = 0,07$. Оценка качества кластеризации с помощью

метрики *силуэт* [11] (выборочное значение 0,47) соответствует плотным кластерам, проекции которых на двумерное пространство представляют собой компактные „облака“ точек (рис. 2).

Дальнейший анализ кластеров позволил объединить их в четыре базовых класса, допускающих очевидную предметную интерпретацию:

— класс I: преимущественно пожилые женщины с высокими уровнем АД и риском сердечно-сосудистых осложнений (вследствие наличия дислипидемии и ишемической болезни сердца, ИБС). Кластеры 1, 2 относятся к разным возрастным группам;

— класс II: пациенты с тяжелой АГ (высокое АД) и нарушениями углеводного и липидного обменов. Кластеры 3, 4 — разные возрастные группы и уровень АД;

— класс III: преимущественно молодые мужчины с отягощенной наследственностью и факторами риска, с поражениями органов-мишеней. Кластер 5;

— класс IV: пациенты с мягкой АГ (АГ 1-й степени, невысокое АД) с факторами риска прогрессирования атеросклероза. Кластеры 6—8 — разные возрастные группы и уровень АД.

Следует отметить, что с течением времени и по мере развития заболевания пациенты могут переходить из кластера в кластер. На рис. 3 приведена диаграмма вероятностных переходов между кластерами. Видно, что ряд кластеров является несовместным (например, в силу гендерных особенностей). При этом каждый кластер связан с риском возникновения сопутствующих заболеваний, непосредственно обусловленных АГ. В табл. 1 представлены выборочные частоты развития сопутствующих патологий, а именно: диабет, нарушение толерантности к глюкозе (НТГ), дислипидемия, гипертрофия левого желудочка (ГЛЖ), хроническая сердечная недостаточность (ХСН), микроальбуминурия, ИБС, хроническая болезнь почек (ХБП), которые далее интерпретируются в модели как условные вероятности.

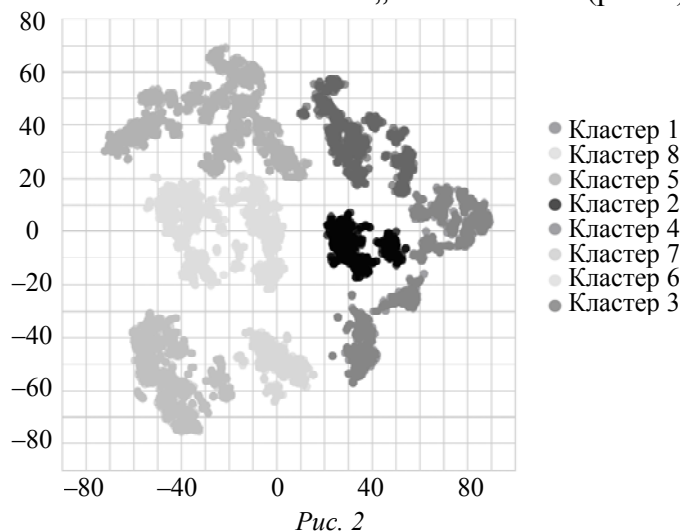


Рис. 2

Таблица 1

№ кластера	Условная вероятность, %												
	Сопутствующие заболевания								Сопутствующие классы препаратов				
	Диабет	НТГ	Дислипидемия	ГЛЖ	ХСН	Микроальбуминурия	ИБС	ХБП	Антиагреганты	Статины	Антикоагулянты	Антиаритмики	Витамины, макро- и микроэлементы
1	20	6	64	0,3	70	0	66	0,3	69	62	8	11	10
2	14	6	66	1	67	0	58	1	66	59	6	11	11
3	13	7	62	2	58	0	50	1	54	49	7	9	6
4	19	6	70	3	72	0	57	0,6	68	56	7	10	7
5	9	5	48	1	54	0,3	34	0,1	37	36	6	7	9
6	14	8	61	2	71	0	58	0,2	62	56	11	12	8
7	17	7	67	1	77	0	64	0,1	72	66	8	14	14
8	15	6	68	0	73	0	56	0	65	57	8	13	9

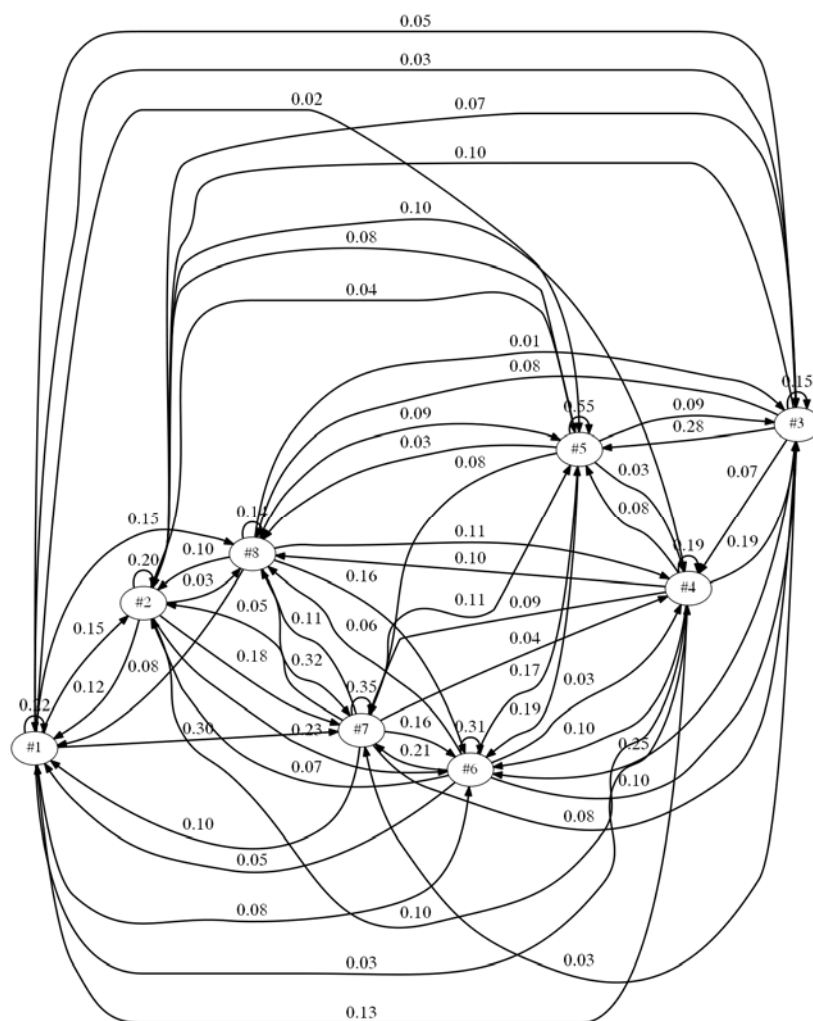


Рис. 3

Поскольку сами кластеры считаются однородными, для пациентов из каждого кластера можно построить модель изменчивости количественных характеристик, например, совместного распределения САД и ДАД в зависимости от возраста, пола и ИМТ. Для приближения условного распределения САД применено трехпараметрическое распределение Вейбулла (где параметр положения отсчитывается от нормативной верхней границы, 140 мм рт. ст.). Для распределения ДАД может быть использовано как распределение Вейбулла, так и нормальное распределение (поскольку оно более симметрично). При этом сами параметры таких распределений являются функциями возраста и ИМТ, оцениваемыми посредством регрессионного анализа.

Для каждого пациента с АГ можно выстроить тактику и стратегию лечения, учитывая рекомендации европейского кардиологического сообщества (*European Society of Cardiology/European Society of Hypertension, (ESH/ESC)*). Однако в отличие от популяционной модели, описывающей все возможные варианты развития заболевания, в рассматриваемом случае целью является назначение *эффективной* антигипертензивной (АГП) терапии на индивидуальном уровне. Критерий эффективности назначенной АГП-терапии — максимизация вероятности достижения целевого уровня АД (140×90 мм рт. ст.) на фоне приема АГП-препаратов в течение не менее 1 месяца, что отражает факт „накопления“ действия препаратов.

На основе этого критерия для каждого класса АГП-препарата был построен классификатор, прогнозирующий его эффективность (метка класса 1) или неэффективность (метка класса 0) для пациента по имеющимся до начала терапии компонентам его цифрового профиля. В основе классификатора — деревья решений на *CART*-алгоритме [12]. Настройка пара-

метров деревьев решений проводилась на основе максимизации показателя доли верно классифицированных объектов (*accuracy*) на тестовой выборке. Качество работы модели характеризуется чувствительностью (*sensitivity*), показывающей способность построенных деревьев решений идентифицировать эффективную терапию, и специфичностью (*specificity*), оценивающей способность моделей определять неэффективное лечение. Достигнутые значения этих показателей представлены в табл. 2. В ходе настройки и валидации модели выявлено, что наиболее важными факторами, влияющими на эффективность АГП-терапии, являются уровень САД (*индекс Джини* = 0,36), возраст (0,29), ИМТ (0,18), наследственность (0,05), дислипидемия (0,08), ХСН (0,04). На основе классификатора для каждого пациента может быть построен сценарий лечения разными классами препаратов, оптимального выбора порядка их применения с учетом совместимости с ранее назначенной терапией.

Таблица 2

Оптимальная глубина дерева; метрики качества	Бета-блокаторы	Ингибиторы АПФ	Антагонисты рецепторов к ангиотензину II	Антагонисты кальция	Диуретики
<i>Применение индекса Джини в качестве критерия разделения предикторов</i>					
Глубина дерева	4	4	4	4	5
Accuracy	0,63	0,58	0,59	0,56	0,5
Sensitivity	0,62	0,12	0,19	0,33	0,25
Specificity	0,56	0,9	0,92	0,71	0,6
<i>Применение энтропийного критерия разделения предикторов</i>					
Глубина дерева	8	8	7	8	5
Accuracy	0,59	0,61	0,61	0,51	0,5
Sensitivity	0,46	0,59	0,3	0,33	0,5
Specificity	0,75	0,56	0,85	0,52	0,4

Таким образом, разработанные методы и модели могут быть заложены в основу технологии популяционного моделирования заболеваемости, развития и лечения АГ. На первом шаге моделируется демографическая обстановка в выбранном регионе (например, с прогнозом на заданное число лет). Далее с помощью метода Монте-Карло формируется начальная популяция пациентов с АГ, которая разбивается пропорционально выявленным выше кластерам. Для пациентов внутри каждого кластера моделируются количественные параметры (САД, ДАД, ИМТ, сопутствующие диагнозы). Затем воспроизводится эволюция самой популяции, которая управляется тремя факторами: обращениями новых пациентов, убылью пациентов преклонного возраста вследствие смертности, а также переходом пациентов из кластера в кластер по мере развития заболевания. При этом на основе представленных выше условных вероятностей разыгрываются случаи регистрации сопутствующих диагнозов. На основе синтетической популяции цифровых профилей пациентов с АГ оцениваются характеристики эффективности лечения согласно различным стратегиям.

Разработанная технология моделирования реализована в виде программного кода на языке Python. Она может быть использована для прогнозирования нагрузки и оптимизации процессов системы здравоохранения (в первую очередь, для оказания амбулаторной помощи и наблюдения), для оценки эффективности внедрения новых препаратов и стратегий лечения, для моделирования сценариев введения новых клинических рекомендаций.

Работа выполнена при финансовой поддержке Российского научного фонда, соглашение № 14-11-00823 (15.07.2014).

СПИСОК ЛИТЕРАТУРЫ

1. Polak S. and Mendyk A. Artificial neural networks based Internet hypertension prediction tool development and validation // Appl. Soft Comput. 2008. Vol. 8, N 1. P. 734—739.
2. Teixeira P. L. et al. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals // J. Am. Med. Informatics Assoc. 2017. Vol. 24, N 1. P. 162—171.
3. Echowffo-Tcheugui J. B., Batty G. D., Kivimä Ki M., and Kengne A. P. Risk Models to Predict Hypertension: A Systematic Review // PLoS ONE. 2013. Vol. 8, N 7. P. e67370. DOI.org/10.1371/journal.pone.0067370
4. Farran B., Channanath A. M., Behbehani K., and Thanaraj T. A. Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait — a cohort study // BMJ Open. 2013. Vol. 3, N 5. P. e002457.
5. Khalilia M., Chakraborty S., and Popescu M. Predicting disease risks from highly imbalanced data using random forest // BMC Med. Inform. Decis. Mak. 2011. Vol. 11, N 1. P. 51.
6. Mancia G. et al. 2013 ESH/ESC guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC) // Eur. Heart J. 2013. Vol. 34, N 28. P. 2159—2219.
7. Friedman C. Semantic Text Parsing for Patient Records // Med. Informatics Knowl. Manag. Data Min. Biomed. 2005. P. 423—448.
8. Бойков И. В., Логинов А. Ю. Демографическая модель с распределенными параметрами // Изв. вузов. Поволжский регион. Физико-математические науки. 2008. № 2. С. 15—20.
9. Lewington S., Clarke R., Qizilbash N., Peto R., and Collins R. Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies // The Lancet. 2002. Vol. 360, N 9349. P. 1903—1913.
10. Van Der Maaten L. and Hinton G. Visualizing Data using t-SNE // J. Mach. Learn. Res. 2008. Vol. 9. P. 2579—2605.
11. de Amorim R. C. and Hennig C. Recovering the number of clusters in data sets with noise features using feature rescaling factors // Inf. Sci. (Ny). 2015. Vol. 324. P. 126—145.
12. Semakova A., Zvartau N., Bochenina K., and Konradi A. Towards Identifying of Effective Personalized Antihypertensive Treatment Rules from Electronic Health Records Data Using Classification Methods: Initial Model // Procedia Comput. Sci. 2017. Vol. 121. P. 852—858.

Сведения об авторах

- Анна Александровна Семакова** — аспирант; Университет ИТМО; кафедра высокопроизводительных вычислений; E-mail: a.a.semakova@gmail.com
- Надежда Эдвиновна Звартау** — канд. мед. наук; Национальный медицинский исследовательский центр имени В. А. Алмазова, организационно-методическое управление по кардиологии и сердечно-сосудистой хирургии; начальник; Университет ИТМО; E-mail: zvartau@almazovcentre.ru
- Сергей Валерьевич Ковальчук** — канд. техн. наук; Университет ИТМО; кафедра высокопроизводительных вычислений; E-mail: sergey.v.kovalchuk@gmail.com
- Александр Валерьевич Бухановский** — д-р техн. наук; Университет ИТМО; мегафакультет трансляционных информационных технологий; директор; E-mail: avb_mail@mail.ru

Поступила в редакцию
15.08.18 г.

Ссылка для цитирования: Семакова А. А., Звартау Н. Э., Ковальчук С. В., Бухановский А. В. Многомасштабное популяционное моделирование процессов развития и лечения артериальной гипертензии // Изв. вузов. Приборостроение. 2018. Т. 61, № 10. С. 922—929.

MULTI-SCALE POPULATION MODELING OF DEVELOPMENT AND TREATMENT OF ARTERIAL HYPERTENSION

A. A. Semakova¹, N. E. Zvartau^{1,2}, S. V. Kovalchuk¹, A. V. Boukhanovsky¹¹ITMO University, 197101, St. Petersburg, Russia
E-mail: a.a.semakova@gmail.com²Almazov National Medical Research Centre, 197341, St. Petersburg, Russia

Arterial hypertension (AH) manifestation and development is associated with many endogenous and exogenous factors. A dynamic population model of AH development in different time scales with consideration of personalized treatment strategies is presented. The overall dynamics of the hypertensive patient group using a demographic model is created. Simulation of AH development at the individual level, is performed with the use of clustering the patients array based on data stored in electronic health records. For patients from each group, models of variability of the quantitative characteristics of the patient's digital profile were developed and the probabilities of the development of concomitant pathologies were estimated. A model is constructed of antihypertensive treatment tactics and strategy using a classifier predicting the most effective class of drugs for the patient based on his/her digital profile individual components collected before the treatment initiation. The developed simulation technology can be used to optimize the health care system processes, primarily for outpatient care, to estimate the effectiveness of new drugs and treatment strategies introduction, and to simulate scenarios for new clinical guidelines introduction.

Keywords: arterial hypertension, dynamic population model, digital patient profile, clustering, CART-algorithm

REFERENCES

1. Polak S. and Mendyk A. *Appl. Soft Comput.*, 2008, no. 1(8), pp. 734–739.
2. Teixeira P.L. et al. *J. Am. Med. Informatics Assoc.*, 2017, no. 1(24), pp. 162–171.
3. Echouffo-Tcheugui J.B., Batty G.D., Kivimä Ki M., and Kengne A.P. *PLoS ONE*, 2013, no. 7(8), pp. e67370. DOI.org/10.1371/journal.pone.0067370
4. Farran B., Channanath A.M., Behbehani K., and Thanaraj T.A. *BMJ Open*, 2013, no. 5(3), pp. e002457.
5. Khalilia M., Chakraborty S., and Popescu M. *BMC Med. Inform. Decis. Mak.*, 2011, no. 1(11), pp. 51.
6. Mancia G. et al. *Eur. Heart J.*, 2013, no. 28(34), pp. 2159–2219.
7. Friedman C. *Med. Informatics Knowl. Manag. Data Min. Biomed.*, 2005, pp. 423–448.
8. Boykov I.V., Loginov A.Yu. *University proceedings. Volga region. Physical and mathematical sciences*, 2008, no. 2, pp. 15–20. (in Russ.)
9. Lewington S., Clarke R., Qizilbash N., Peto R., and Collins R. *The Lancet*, 2002, no. 9349(360), pp. 1903–1913.
10. Van Der Maaten L. and Hinton G. J. *Mach. Learn. Res.*, 2008, no. 9, pp. 2579–2605.
11. de Amorim R.C. and Hennig C. *Inf. Sci. (Ny)*, 2015, no. 324, pp. 126–145.
12. Semakova A., Zvartau N., Bochenina K., and Konradi A. *Procedia Comput. Sci.*, 2017, no. 121, pp. 852–858.

Data on authors

- Anna A. Semakova** — Post-Graduate Student; ITMO University, Department of High-Performance Computing; E-mail: a.a.semakova@gmail.com
- Nadezhda E. Zvartau** — PhD; Almazov National Medical Research Centre, Organizational Department on Cardiology and Cardiovascular Surgery; Head of the Department; ITMO University, E-mail: zvartau@almazovcentre.ru
- Sergey V. Kovalchuk** — PhD; ITMO University, Department of High-Performance Computing; E-mail: sergey.v.kovalchuk@gmail.com
- Alexander V. Boukhanovsky** — Dr. Sci., Professor; ITMO University, School of Translational Information Technologies; Head of the School; E-mail: avb_mail@mail.ru

For citation: Semakova A. A., Zvartau N. E., Kovalchuk S. V., Boukhanovsky A. V. Multi-scale population modeling of development and treatment of arterial hypertension. *Journal of Instrument Engineering*. 2018. Vol. 61, N 10. P. 922–929 (in Russian).

DOI: 10.17586/0021-3454-2018-61-10-922-929