

DOI 10.37882/2223–2966.2022.06.29

РАЗРАБОТКА КЛАССИФИКАТОРА НАГРУЗКИ ДЛЯ МОДЕЛИ ОЦЕНКИ СОСТОЯНИЯ ОПЕРАТОРА КИБЕРФИЗИЧЕСКОЙ СИСТЕМЫ

DEVELOPMENT OF A LOAD CLASSIFIER FOR A MODEL FOR ASSESSING THE STATE OF A CYBERPHYSICAL SYSTEM OPERATOR

A. Penin

Summary. During the research work, the main biological markers of the human body were studied and those of them that met the requirements were selected for further research. A classifier of the employee's workload level was developed based on the decision tree method, the classification accuracy was 76%, the value of the loss function was 0.23. In the future, the classifier of the employee's workload level and biological markers will be combined into a model for assessing the state of the operator of a cyber-physical system.

Keywords: neural networks, multi-class classification, biomarkers, models, systems, state.

Пенин Андрей Семенович

ФГАОУ ВО «Национальный исследовательский университет ИТМО», Санкт-Петербург
gorogtheghost@gmail.com

Аннотация. В ходе исследовательской работы были изучены основные биологические маркеры человеческого организма и отобраны для дальнейших исследований те из них, которые удовлетворяли поставленным требованиям. Был разработан классификатор уровня нагрузки сотрудника на основе метода дерева решений, точность классификации составила 76%, значение функции потерь составило 0,23. В дальнейшем классификатор уровня нагрузки сотрудника и биологические маркеры будут объединены в модель оценки состояния оператора киберфизической системы.

Ключевые слова: нейросети, мультиклассовая классификация, биомаркеры, модели, системы, состояние.

Человек был и остается крайне уязвимым элементом любой системы, частью которой он является, будь то с физической или информационной точки зрения. И если раньше информационные технологии позволяли повысить его безопасность лишь в цифровом пространстве, современное их развитие дает возможность выйти за его пределы, особенно в ситуациях, когда люди особенно уязвимы, например, при работе в условиях Крайнего Севера.

Работники удаленных предприятий, работающие вахтовым методом, подвергаются постоянной опасности природного и техногенного характера [4]. С целью обеспечения их личной безопасности и безопасности условий труда, а также для контроля за их деятельностью возможно использовать носимые устройства (таких как «умные часы» и пр.), которые способны собирать и анализировать общую информацию о состоянии организма по так называемым «биомаркерам» [1].

Под «биомаркерами» при этом следует понимать набор биологических признаков, которые в совокупности могут быть использованы для идентификации текущего состояния организма [2]. К таковым признакам можно отнести:

- ◆ температуру тела;
- ◆ сатурацию;
- ◆ кровяное давление;

- ◆ пульс;
- ◆ и др.

Все указанные характеристики относятся к биомаркерам нулевого типа, т.е. указывающим на наличие заболеваний и связанных с ними клинических проявлений. Используя собранные с помощью носимых устройств статистические данные [3], а также современные технологии анализа данных с помощью нейросетей становится возможным решить следующие задачи:

- ◆ оценить текущие физиологические процессы в организме;
- ◆ спрогнозировать риск появления новых и развития старых заболеваний;
- ◆ оценить эффективность лечения текущих заболеваний;
- ◆ на основании собранных данных определить негативные факторы окружающей среды;
- ◆ сформировать градацию сложности выполняемых на предприятии физических работ.

Решение указанных задач позволяет добиться сразу нескольких целей. В первую очередь такое решение позволит сформировать и поддерживать в постоянно актуальном состоянии базу данных о медицинском состоянии работников, работающих в неблагоприятных условиях. Такая база данных может быть использована для проведения аналитической работы по оценке эф-

фективности решения различных категорий задач различными сотрудниками, влиянии различных категорий задач на медицинское состояние работников и др. Немаловажным будет и возможность раннего выявления различных профессиональных заболеваний [5, 9], развитие которых может быть незаметно на раннем этапе без специализированного обследования, но которые могут повлиять на качество жизни работников предприятия.

Следует отметить, что сбор перечисленных данных и их анализ в данный момент уже активно применяется во многих видах деятельности в форме регулярных медицинских осмотров работников, однако такие процедуры имеют высокую (из-за необходимости проводить их для каждого работника в течение длительного периода времени) стоимость и разделены между собой значительным временным промежутком, в ходе которого могут возникнуть опасности для медицинского состояния работника, которые не будут выявлены до следующего регулярного осмотра. Предлагаемое решение предполагает постоянный автоматизированный мониторинг состояния работников предприятия, что должно позволить использовать информационные технологии для снижения риска получения травмы или профессионального заболевания при работе на предприятии [9]. При этом разрабатываемое решение имеет возможность оставаться более экономически эффективным, чем применяемые в данный момент методики.

Таким образом, формализуя уже сказанное, актуальность данного исследования обусловлена необходимостью поддерживать высокий уровень безопасности работников предприятий располагающихся в неблагоприятных условиях и сложностью обеспечения заданного уровня безопасности.

Целью настоящего исследования является разработка классификатора нагрузки для модели оценки состояния оператора киберфизической системы. Данный классификатор основывается на пульсе оператора и применяется для оценки уровня нагрузки. В рамках исследования нагрузка делится на три уровня: низкая-покой, средняя, высокая.

Научная новизна проводимого исследования определяется оригинальностью разрабатываемой системы оценки оператора киберфизической системы. Для этого используется классификатор, основанный на методе «дерева решений». Уникальность решения основывается на способе получения данных (с носимого устройства), а также на последующем применении разрабатываемого классификатора как части многокомпонентной системы. Такой подход позволит, используя биомаркеры и данные о активности оператора, упростить интерпре-

тацию статистических данных, что также является новым оригинальным решением.

Теоретическая значимость исследования заключается в том, что собираемые данные могут быть использованы для изучения влияния факторов окружающей среды и медицинских факторов на работников предприятий, расположенных в условиях Крайнего Севера с возможностью последующего тестирования. Практическая значимость исследования состоит в возможности повышения безопасности предприятий путем внедрения разработанной модели и исследовании результатов её работы.

При подготовке к исследованию была сформирована выборка испытуемых состоящая из мужчин и женщин в возрасте от 27 до 35 лет. В набор данных была собрана информация о поле, росте, весе, склонности к курению, пульсе в состоянии покоя. Общее число испытуемых составило 232 человека. Затем участником было предложено проделать упражнения нескольких степеней интенсивности. Были записаны данные о пульсе участников во время прохождения упражнений. Данные упражнения были промаркированы. Были установлены следующие категории физической активности:

- ◆ легкая физическая активность/состояние покоя — в данную категорию попала деятельность, не выходящая за рамки повседневных задач в условиях нормального уровня стресса;
- ◆ средняя физическая активность — в данную категорию попала деятельность несколько повышающая сердечные сокращения. В качестве примера такой деятельности можно назвать быструю ходьбу, езда на велосипеде по ровной поверхности и т.д.;
- ◆ высокая физическая активность — в эту категорию попала деятельность, которая в значительной степени влияет на сердечные сокращения: бег, тяжелые физические упражнения, плавание на дистанцию и т.п.

При разработке классификатора рассматривались данные некурящих людей обоего пола. Таким образом число испытуемых снизилось до 206 человек. Средний рост испытуемых составил $165,71 \pm 6,31$ см у женщин и $180,34 \pm 6,04$ см у мужчин. Средний вес составил $61,15 \pm 6,98$ кг у женщин и $80,12 \pm 12,64$ кг у мужчин. Из них 32 человека занимались легкой физической активностью, 78 человек занимались физической активностью средней тяжести, 96 человек испытывали тяжелые физические нагрузки.

Для разработки классификатора были использованы следующие методы: логистическая регрессия, дерево принятия решений, случайный лес. В качестве функции

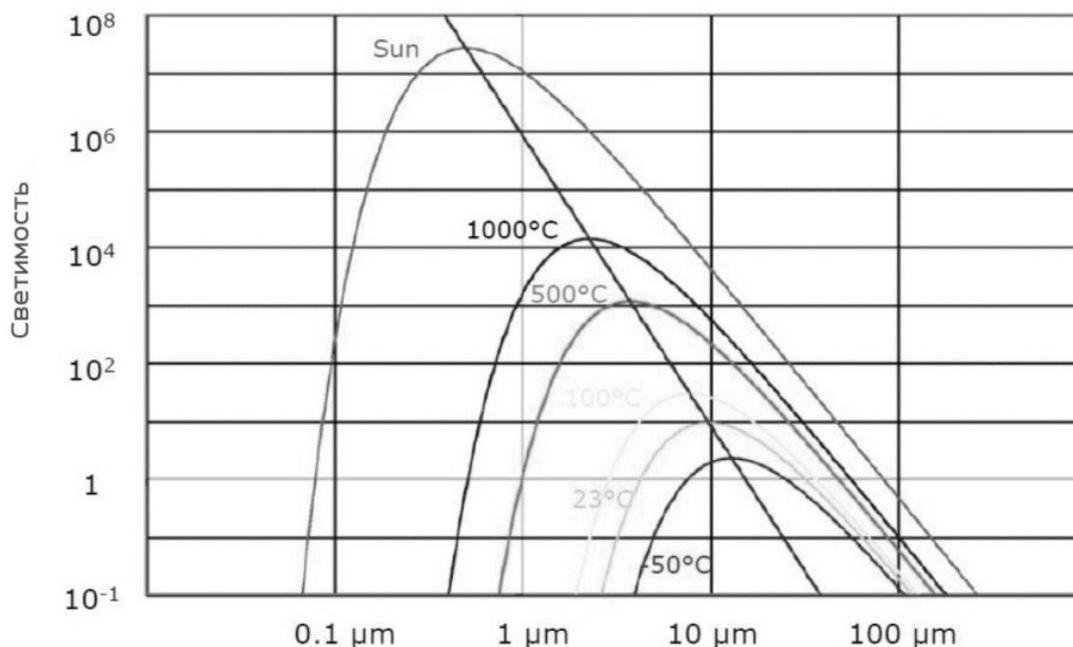


Рис. 1. График температуры объекта в зависимости от исходящего от него ИК излучения

потерь использовалась средняя абсолютная ошибка. В качестве меры оценки качества работы модели использовалась достоверность (accuracy) полученной модели. Размер тестовой выборки составил 20% от всего набора данных.

Кроме того, были сформированы требования к биомаркерам, используемым в дальнейших исследованиях, и отобраны биомаркеры удовлетворяющие выставленным требованиям.

В данном разделе будут подробно рассмотрены использованные в ходе разработки биомаркеры состояния оператора, используемые методы их получения, преимущества и недостатки данных методов. Кроме того, в данном разделе будет приведена информация по исходным данным, используемым при построении будущей модели.

Все отобранные маркеры отвечали следующим требованиям:

- ◆ показательность изменений значений биомаркера. Под показательностью в данном случае имеется ввиду прямая зависимость между изменением биомаркера и изменением состояния работника;
- ◆ возможность простого считывания биомаркеров средствами носимого устройства;
- ◆ экономическая эффективность такого считывания. В данном случае экономически эффективным считается отсутствие необходимости

разработки специализированного носимого устройства и использование базового доступного функционала.

В качестве первого маркера в модели используется температура тела оператора. Данный параметр может изменяться под воздействием внешней среды, физических нагрузок, пищи, стресса, умственной работы и т.п. Однако одним из важнейших факторов изменения температуры тела является ответ иммунитета на заболевание. Рассмотрим вопрос изменения температуры подробнее.

Описанные выше причины можно поделить на внешние (возникшие в результате воздействия на человека извне или его деятельности) и внутренние (возникшие в результате процессов внутри человеческого тела). При этом внутренние изменения температуры тела являются более четким маркером медицинского состояния оператора, однако однозначная классификация источника изменения температуры тела с помощью носимого устройства является отдельной задачей, не решаемой в рамках текущего исследования. Вместе с тем, постоянный контроль температуры операторов позволит заранее определить изменения состояния их здоровья, условий труда или вида деятельности.

Для получения информации о температуре тела с помощью носимого устройства предполагается использование метода с регистрацией инфракрасного(ИК)-излучения, на основании интенсивности кото-

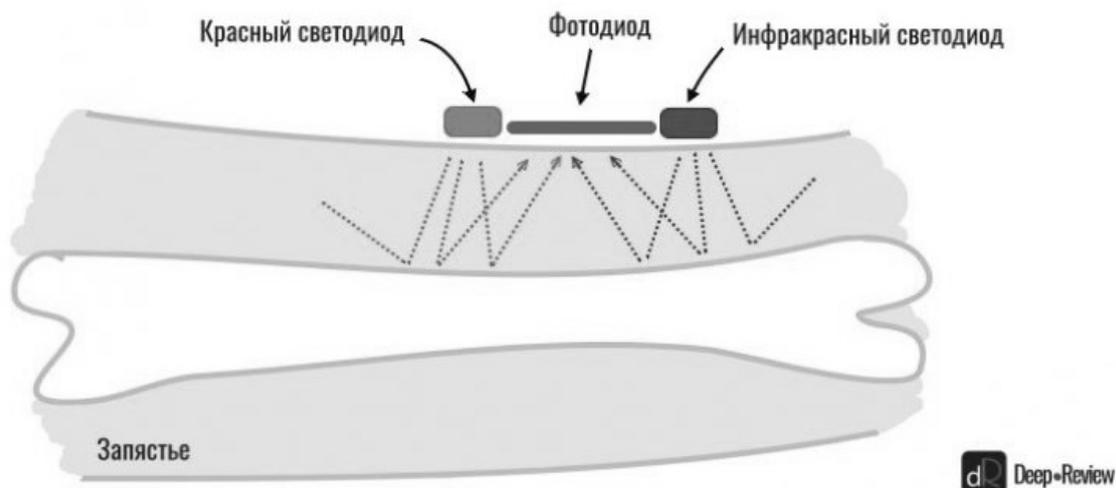


Рис. 2. Схема работы пульсоксиметров основанных на отраженном свете

рого и вычисляется температура тела. Преимуществом данного метода является возможность измерения температуры даже на расстоянии, при этом тепловая инерция ИК-датчиков не превышает аналогичных показателей электронных термометров [11]. Однако у данной технологии существует ряд недостатков, среди которых присутствуют следующие:

- ◆ разность излучения у разных материалов;
- ◆ высокая (до 0.5 градусов) погрешность при неправильной эксплуатации;
- ◆ высокая стоимость в сравнении с традиционными приборами;
- ◆ необходимость в частом техосмотре.

На рисунке 1 представлен график температуры объекта в зависимости от интенсивности исходящего излучения.

Следующим рассматриваемым маркером является сатурация. Сатурацией называют показатель уровня насыщения крови кислородом, при этом чем выше уровень насыщения, тем выше сатурация. Нормальным считается уровень сатурации от 94% [6], при этом к причинам по которым уровень сатурации может измениться относятся:

- ◆ избыточный вес;
- ◆ анемическое состояние;
- ◆ повышенное артериальное давление;
- ◆ дисфункция щитовидной железы;
- ◆ болезни сердца;
- ◆ нарушение работы системы кровообращения;
- ◆ травмы.

Как видно из перечисленного списка, все рассматриваемые причины так или иначе связаны с ухудшением

медицинского состояния работника. Таким образом изменения сатурации является одним из важнейших маркеров для модели, позволяющим с большей точностью определить разницу между изменениями, возникающими в результате физической нагрузки, и изменениями, являющимися следствиями травм.

Для оценки сатурации используется модифицированная схема стандартного пульсоксиметра, реализованная таким образом, чтобы быть помещенной в носимое устройство. Для этого в корпусе устройства рядом с фотодиодом устанавливаются красный и инфракрасный светодиоды, которые испускают свет, отражающийся от кости и улавливаемый фотодиодом. Проходя через сосуды такой свет поглощается гемоглобином, причем молекулы с кислородом и молекулы без кислорода по разному поглощают волны разной длины. Первые лучше поглощают красный свет, вторые — инфракрасный. Таким образом для определения уровня насыщения крови кислородом достаточно узнать, сколько инфракрасного и красного света было поглощено гемоглобином. Схема работы такого устройства приведена на рисунке 2.

Недостатком такого устройства является его неточность в условиях, когда исследование проводится при ярком свете, если объект, на который установлен датчик, находится в движении, если свет датчиков проходит через отражающие поверхности, если сам пульсоксиметр установлен неправильно. Все эти условия могут встречаться на реальном производстве, что необходимо учесть при обработке данных с устройства. Кроме того, в некоторых чрезвычайных ситуациях показания пульсоксиметра могут быть ложными, например при отравлении угарным газом. Однако, не смотря на это,

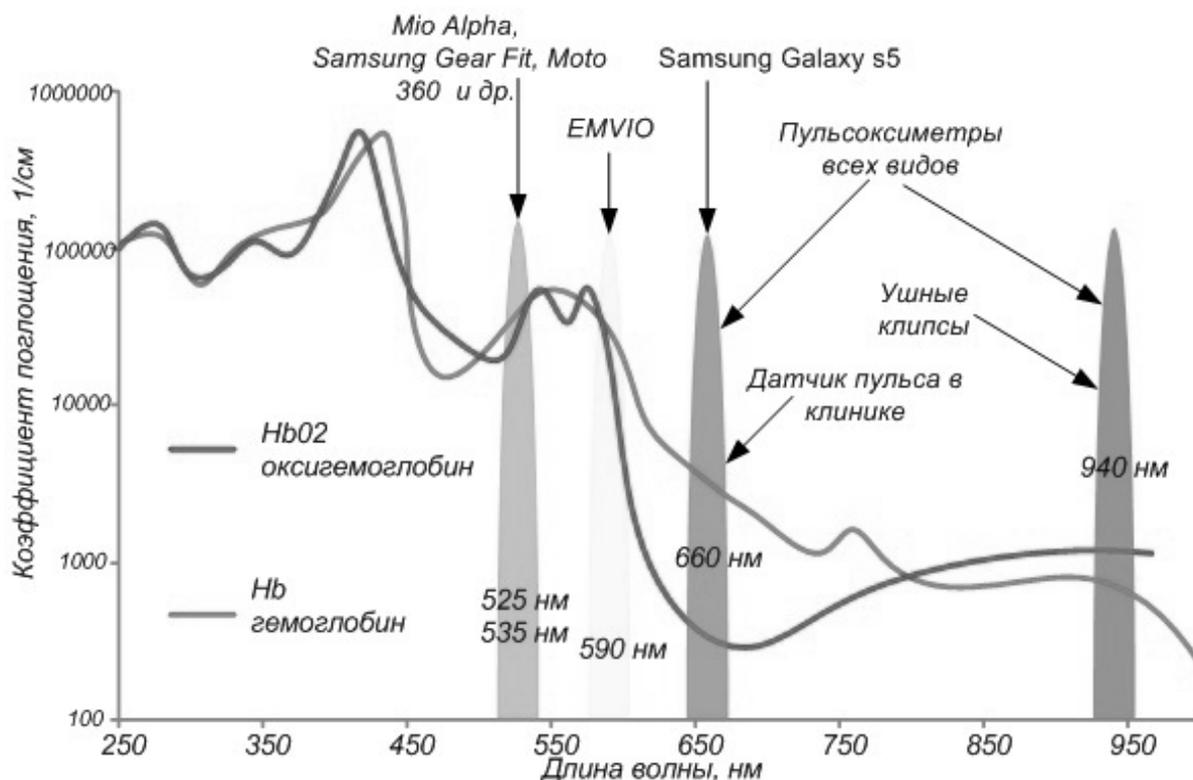


Рис. 3. Кривая поглощения света гемоглобином с указанием основных используемых спектров излучения фотоплетизмографических датчиков и устройств, в которых они применяются

использование сатурации как одного из определяющих маркеров в совокупности является оправданным.

Ещё одним биомаркером, используемом в разработанной модели, является пульс. Пульс — это толчкообразные колебания стенок артерий, связанные с сердечными циклами. В более широком смысле под пульсом понимается любое изменение в сосудистой системе, связанное с деятельностью сердца. При этом средними нормальными показателями считается частота пульса от 60 до 100 ударов в минуту [7]. При этом следует отметить, что характеристика артериального пульса (т.е. полученного путем оценки количества сокращений стенок кровеносных сосудов) и характеристика «сердечного» пульса (измеряемого по электрической активности сердца) не совпадают между собой. Они имеют разную физиологическую природу, разную форму временной кривой, разный фазовый сдвиг и, как следствие, требуют разных методов регистрации и обработки. В данном исследовании использовалась характеристика артериального пульса.

При этом следует понимать, что изменения пульса сами по себе не могут быть средствами идентификации состояния носителя. Повышение пульса может быть связано с употреблением кофеин-содержащих продуктов,

физическая нагрузка, применение некоторых препаратов, заболеваниями сердечно-сосудистой системы, болезни эндокринной системы. К понижению пульса, в свою очередь, может привести прием медицинских препаратов, ряд инфекционных заболеваний, вегетативная патология нервной системы и др. Таким образом данные по пульсу несут в себе в первую очередь уточняющий характер по отношению к остальным, а сам пульс является вспомогательной характеристикой.

Для измерения пульса в носимых устройствах используется оптический метод (фотоплетизмография). В нем сужение и расширение сосудов под действием артериальной пульсации вызывает соответствующие изменения амплитуды сигнала, получаемого фотоприемником. При этом важную роль играет длина волны сигнала. На сигналы с разной длиной по-разному влияют различные артефакты, такие как: артефакты движения, индивидуальные свойства кожи, толщина подкожного слоя жира и прочее. На рисунке 3 представлена кривая поглощения света гемоглобином и основные спектры излучения фотоплетизмографических датчиков.

Как видно из рисунка, оптимальными для решения задачи определения пульса являются значения длины

волны в промежутке 500–600 нм. При этом чаще всего в современных носимых устройствах используются длины волны 525–535 нм (зеленый свет).

Последним из использованных в разработке биомаркеров стало давление. Кровяное давление — это то давление, которое кровь оказывает на стенки сосудов. Данный маркер является параметром, характеризующим работу кровеносной системы, и определяется объемом крови, перекачиваемым за единицу времени сердцем и сопротивлением сосудистого русла. В данной исследовательской работе под давлением подразумевается в артериальное давление, формирующееся из пары значений: систолического давления (давления в момент сжатия сердца) и диастолического давления (давления в момент расслабления сердца). Давление может меняться вследствие смены времени суток, погоды, психологического состояния человека, приема препаратов или медикаментов, питания, физических нагрузок [8]. С его помощью удобно определять общую нагрузку на организм во время проведения работ и ответ организма на эту нагрузку. Таким образом открывается возможность на формирование более оптимальных условий труда.

Существует несколько методик по оценке давления с помощью носимых устройств, ниже приведем три наиболее распространенных:

- ◆ комбинированный с ЭКГ датчиком. Данная методика позволяет получить достаточно точный результат. Однако она сложна в реализации из-за необходимости наличия ЭКГ устройства;
- ◆ осциллометрический метод, применяемый также в медицинских тонометрах. Является наиболее точным из отобранных, одна в то же время наиболее трудных в реализации;
- ◆ расчет на основании данных с пульсометра и внешних пользователем данных о росте, весе, возрасте, поле и т.д. Результаты полученные таким образом наименее точны и эффективны, но при этом такой вариант является самым простым в реализации и доступным. Используемые при разработке модели данные были получены именно таким образом.

Датчики для считывания отобранных биомаркеров будут размещены в носимых устройствах в виде браслетов, расположенных на левой и правой руках оператора системы. В качестве итогового получаемого значения данных от оператора будет использоваться среднее значение между показателями датчиков на обеих руках.

Одной из систем, разрабатываемых в рамках анализа получаемых таким образом данных является классификатор нагрузки, основанный на данных о пульсе.

В качестве метода для поиска решений для классификатора на основе логистической регрессии был использован метод `liblinear`, лучшим образом подходящий для небольших наборов данных. Норма регуляризации выбрана l_2 , параметр регуляризации $C=1$. Так как задача подразумевает мультиклассовую классификацию, был также задействован параметр `multi_class` в значении «ovr». Средняя взвешенная достоверность модели составила 0,73. Значение средней абсолютной ошибки составило 0,26.

При работе с деревом решений использовался критерий Джини [10]. Размер максимальной глубины дерева был выбран 6, минимальное количество листьев составило 10. Данные параметры были подобраны эмпирическим путем для повышения показателей итогового классификатора. В результате средняя взвешенная достоверность модели составила 0,76, значение средней абсолютной ошибки составило 0,23.

В работе с случайным лесом также был задействован критерий Джини. Количество составляемых деревьев (`n_estimators`) было определено как 10, в качестве максимальной глубины было установлено значение 5. Данные параметры также были подобраны эмпирически. Результатом работы классификатора стала средняя взвешенная достоверность модели 0,71 и значение средней абсолютной ошибки 0,28.

По итогу сравнения результатов работы классификаторов можно сделать вывод, что достоверность результатов при использовании дерева решений является наибольшей, а значение средней абсолютной ошибки — наименьшим.

Обсуждение

Результатом данного исследования стало создание классификатора уровня нагрузки оператора киберфизической системы на основании данных о пульсе, получаемых с носимого устройства оператора. Данный классификатор основывается на методе дерева решений. Значимость полученного результата состоит в том, что полученный классификатор позволит в дальнейшем определять уровень физической нагрузки, выполняемой работником в момент времени, что в сочетании с иными биомаркерами, рассмотренными в данном исследовании, позволит сформировать модель состояния оператора киберфизической системы в каждый момент времени.

Достоверность результатов работы полученного классификатора составляет 76%. Данный результат является приемлемым в рамках текущего научного исследования, однако может быть улучшен при дальнейшей работе над

разрабатываемой моделью. Причиной получения такого результата может заключаться в малом объеме рассматриваемой выборки. Данная проблема может быть решена при внедрении результатов разработки на реальное предприятие, где будет иметься возможность получения большого количества данных. Кроме того, существует возможность, что лучший результат мог быть достигнут с использованием иных методов классификации, например, с помощью нейронных сетей. Данный вопрос также может быть исследован в дальнейшем.

В ходе проведения исследования были изучены различные биомаркеры и отобраны те, которые из них, которые удовлетворяют требованиям, выставляемым целями исследования. Итоговыми биомаркерами стали: температура тела, сатурация, пульс, артериальное давление. В качестве маркеров деятельности были использованы показания акселерометра и гироскопа, расположенных в браслетах на правой и левой руке работника. Теоретическая значимость данного исследования заключается в формировании возможности изучения влияния окружающей и рабочей среды на сотрудников предприятия в условиях вредоносного производства. Собираемые с помощью носимых устройств статистические данные будут иметь меньшую точность, чем те, которые были получены в результате специальных обследований, однако они будут собираться в течение большего периода времени и охватывать различные виды деятельности сотрудников предприятия. В будущем будут проводиться

дальнейшие исследования на данную тематику. Практическая значимость состоит в формировании набора требований к оборудованию, достаточному для получения приемлемой информации о медицинском состоянии работников предприятия.

Был разработан классификатор уровня физической нагрузки работника на базе дерева решений. Данный классификатор позволяет со средней точностью (76%) определить величину физической нагрузки, испытываемой оператором системы. Практическая значимость данного исследования состоит в потенциальной возможности контроля деятельности работников предприятия.

Дальнейшие исследования будут направлены на повышение точности работы классификатора, сборке большей выборки данных, объединение разработанного классификатора нагрузки и системы считывания биомаркеров в единую информационную модель оценки состояния оператора киберфизической системы. Такая модель позволит использовать технологии искусственного интеллекта для определения медицинского состояния работника на удаленном предприятии, учитывая при этом факторы окружающей среды и его непосредственной деятельности. Таким образом информационные технологии позволят добиться улучшения качества жизни операторов и повышения их безопасности при работе даже в самых неблагоприятных условиях.

ЛИТЕРАТУРА

1. Биомаркеры — индикаторы состояния здоровья [Электронный ресурс]. — Режим доступа: <https://medinteres.ru/interesnyie-faktyi/biomarkeryi.html> (дата обращения 28.03.2022)
2. Долгов В.В., Шевченко О.П., Шевченко А.О. Биомаркеры в лабораторной диагностике // Биомаркеры в лабораторной деятельности — 2013 — с. 14
3. Интернет вещей и его значение для промышленности [Электронный ресурс] — Режим доступа: https://www.pwc.ru/ru/publications/iot-for-industry_.pdf (дата обращения 28.03.2022)
4. Корнеева Я.А., Дубинина Н.И., Симонова Н.Н., Дегтева Г.Н., Федотов Д.М. Риски в профессиональной деятельности вахтовых работников в условиях Крайнего Севера // Бюллетень ВСНЦ СО РАМН — 2013 — № 3(91) Часть 2 — с. 83
5. Литвяков А.М., Щупакова А.Н. Общие вопросы профессиональной патологии // Профессиональные заболевания — 2011 — с. 7
6. О сатурации кислорода в крови [Электронный ресурс]. — Режим доступа: <https://aptstore.ru/articles/saturatsiya-kisloroda-v-krovi/> (дата обращения 28.03.2022)
7. Пульсовое давление в крови [Электронный ресурс]. — Режим доступа: <https://cyberleninka.ru/article/n/pulsovoe-davlenie-krovi-rol-v-gemodinamike-i-prikladnye-vozmozhnosti-v-funktsionalnoy-diagnostike> (дата обращения 28.03.2022)
8. Повышенное давление: причины и особенности лечения [Электронный ресурс]. — Режим доступа: <https://aptstore.ru/articles/povyshennoe-davlenie-prichiny-i-osobennosti-lecheniya/> (дата обращения 28.03.2022)
9. Стасева Е.В., Теплякова Н.А. Медосмотры как факторы ранней диагностики профзаболеваний // Молодой ученый — 2016 — № 18.1(122.1) — с. 21–23
10. Соколов Е.А. Решающие деревья // Лекции ФКН ВШЭ — 2018 — с. 5.
11. Тепловая инерция температурных датчиков [Электронный ресурс]. — Режим доступа: <https://isup.ru/articles/16/15436/> (дата обращения 28.03.2022)