

## ПРИБОРЫ, СИСТЕМЫ И ИЗДЕЛИЯ МЕДИЦИНСКОГО НАЗНАЧЕНИЯ

УДК 004.891

### ВИРТУАЛЬНЫЕ ПОТОКИ В ГИБРИДНЫХ РЕШАЮЩИХ МОДУЛЯХ КЛАССИФИКАЦИИ СЛОЖНОСТРУКТУРИРУЕМЫХ ДАННЫХ

*Статья поступила в редакцию 18.10.2018, в окончательном варианте – 12.11.2018.*

**Киселев Алексей Викторович**, Юго-Западный государственный университет, 305004, Российская Федерация, г. Курск, ул. Челюскинцев, 19, корпус Б, преподаватель кафедры информационных систем и технологий, ORCID <https://orcid.org/0000-0001-7228-0281>, SCOPUS <https://www.scopus.com/authid/detail.uri?authorId=%206603139063>, Researcher ID <http://www.researcherid.com/rid/S-9914-2018>, [https://elibrary.ru/author\\_profile.asp?id=97559](https://elibrary.ru/author_profile.asp?id=97559), e-mail: Kisevalexey1990@gmail.com

**Савинов Денис Юрьевич**, Юго-Западный государственный университет, 305004, Российская Федерация, г. Курск, ул. Челюскинцев, 19, корпус Б, аспирант, e-mail: marina-savinova-93@mail.ru

**Филист Сергей Алексеевич**, Юго-Западный государственный университет, 305004, Российская Федерация, г. Курск, ул. Челюскинцев, 19, корпус Б, доктор технических наук, профессор, ORCID <https://orcid.org/0000-0003-1358-671X>, SCOPUS <https://www.scopus.com/authid/detail.uri?authorId=6603139063>, Researcher ID <http://www.researcherid.com/rid/O-4610-2015>, e-mail: SFilist@gmail.com, [https://elibrary.ru/author\\_profile.asp?id=251980](https://elibrary.ru/author_profile.asp?id=251980)

**Шаталова Ольга Владимировна**, Юго-Западный государственный университет, 305004, г. Курск, ул. Челюскинцев, 19, корпус Б,

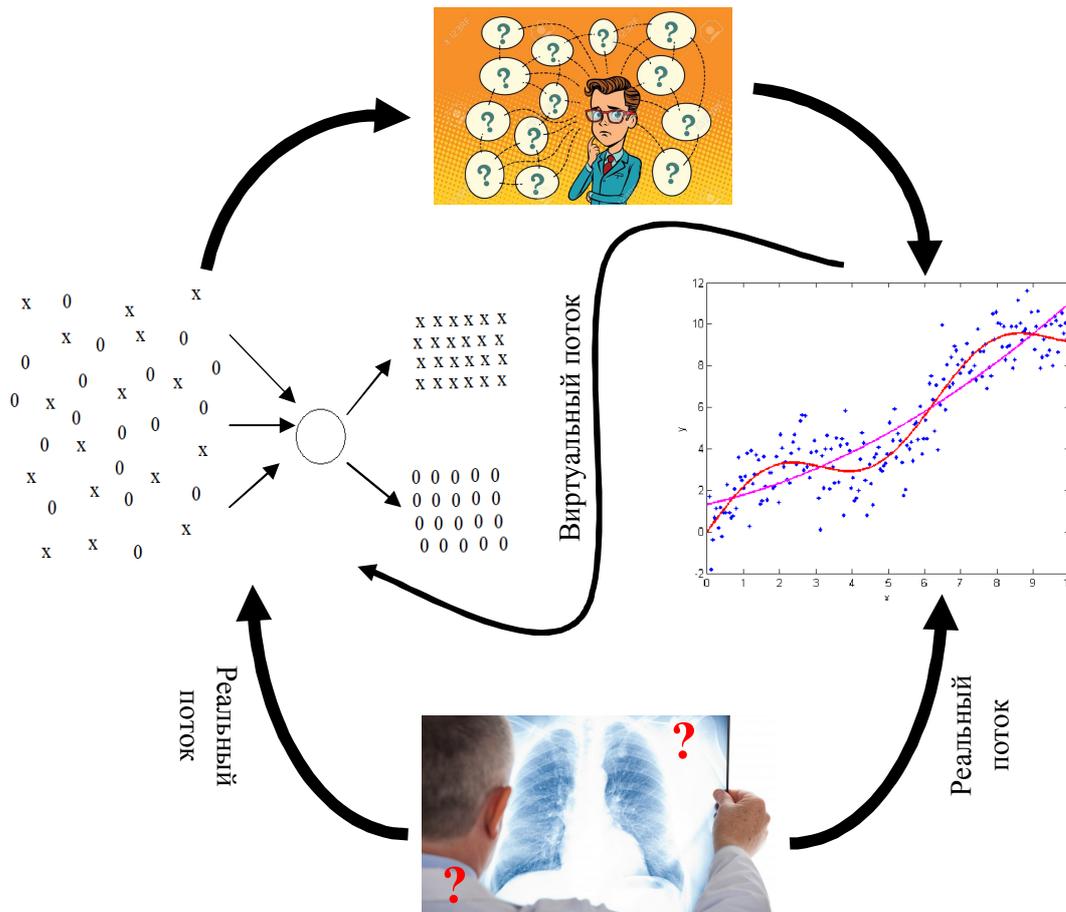
кандидат технических наук, доцент, ORCID <https://orcid.org/0000-0002-0901-9272>, SCOPUS <https://www.scopus.com/authid/detail.uri?authorId=24477712800>, Researcher ID <http://www.researcherid.com/rid/C-3687-2015>, e-mail: shatolg@mail.ru, [https://elibrary.ru/author\\_profile.asp?id=673680](https://elibrary.ru/author_profile.asp?id=673680)

**Жилин Валерий Вячеславович**, Курский институт кооперации, филиал Белгородского университета потребительской кооперации, экономики и права, 305004, Российская Федерация, г. Курск, ул. Радищева, 116, кандидат технических наук, доцент, e-mail: vzhilin61@gmail.com, [https://elibrary.ru/author\\_profile.asp?id=321882](https://elibrary.ru/author_profile.asp?id=321882)

Для классификации сложноструктурируемых данных предложены гибридные решающие модули с виртуальными потоками. Виртуальные потоки отражают скрытые системные связи между наблюдаемыми параметрами процесса или системы. При этом вектор информативных признаков состоит из двух подвекторов, первый из которых соответствует реальным потокам, а второй – виртуальным потокам. Для формирования виртуального потока разработана обобщенная рекуррентная структурная схема. Схема позволяет формировать гибридный вектор информативных признаков, состоящий из двух подвекторов. Один из них состоит из исходных информативных признаков, а второй – из информативных признаков, полученных на основе моделирования системных связей между информативными признаками первого подвектора. Для формирования второго подвектора используются как системные связи между исходным пространством информативных признаков, так и вновь сформированные латентные переменные. Это позволяет реализовать рекуррентный процесс формирования пространства информативных признаков. Предложенный метод формирования нелинейных моделей виртуальных потоков основан на МГУА-моделировании. В методе для получения моделей влияния реальных потоков на виртуальные потоки используются нейронные сети на нелинейных адалинах. Метод позволяет формировать подвектор латентных переменных неограниченной размерности. На основе предложенного метода сформирована структура гибридной решающей системы с виртуальными потоками, предназначенная для классификации сложноструктурируемых данных. Структура позволяет учесть латентные потоки для информативных признаков (виртуальные потоки) – они определяются на основе статистических и экспертных исследований связей между исходными информативными признаками. В свою очередь это дает возможность агрегировать четкие и нечеткие решающие правила, обеспечивая требуемое качество принимаемых решений при разнородной и плохоформализуемой структуре классов. Проведенное авторами имитационное моделирование показало эффективность метода.

**Ключевые слова:** гибридный, решающая система, латентная переменная, МГУА-модель, нейронная сеть, нечеткая логика принятия решений, агрегаторы нечетких решающих правил

## Графическая аннотация (Graphical annotation)



**VIRTUAL FLOWS IN HYBRID DECISION MODULES  
OF CLASSIFICATION OF COMPLEX-STRUCTURED DATA**

*The article was received by editorial board on 18.10.2018, in the final version – 12.11.2018.*

**Kiselyov Alexey V.**, Southwest State University, building B, 19 Chelyuskintsev St., Kursk, 305004, Russian Federation,

Teacher of Information Systems and Technologies Department, ORCID <https://orcid.org/0000-0001-7228-0281>, SCOPUS <https://www.scopus.com/authid/detail.uri?authorId=%206603139063>, Researcher ID <http://www.researcherid.com/rid/S-9914-2018>, [https://elibrary.ru/author\\_profile.asp?id=97559](https://elibrary.ru/author_profile.asp?id=97559), e-mail: [Kiselevalexey1990@gmail.com](mailto:Kiselevalexey1990@gmail.com)

**Savinov Denis Yu.**, Southwest State University, building B, 19 Chelyuskintsev St., Kursk, 305004, Russian Federation,

post-graduate student, e-mail: [marina-savinova-93@mail.ru](mailto:marina-savinova-93@mail.ru)

**Filist Sergey A.**, Southwest State University, building B, 19 Chelyuskintsev St., Kursk, 305004, Russian Federation,

Doct. Sci. (Engineering), Professor, ORCID <https://orcid.org/0000-0003-1358-671X>, SCOPUS <https://www.scopus.com/authid/detail.uri?authorId=6603139063>, Researcher ID <http://www.researcherid.com/rid/O-4610-2015>, e-mail: [SFilist@gmail.com](mailto:SFilist@gmail.com), [https://elibrary.ru/author\\_profile.asp?id=251980](https://elibrary.ru/author_profile.asp?id=251980)

**Shatalova Olga V.**, Southwest State University, building B, 19 Chelyuskintsev St., Kursk, 305004, Russian Federation,

Cand. Sci. (Engineering), Associate Professor, ORCID <https://orcid.org/0000-0002-0901-9272>, SCOPUS <https://www.scopus.com/authid/detail.uri?authorId=24477712800>, Researcher ID <http://www.researcherid.com/rid/C-3687-2015>, e-mail: [shatolg@mail.ru](mailto:shatolg@mail.ru), [https://elibrary.ru/author\\_profile.asp?id=673680](https://elibrary.ru/author_profile.asp?id=673680)

**Zhilin Valery V.**, Kursk Institute of Cooperation, branch of Belgorod University of Consumer Cooperation, Economics and Law, 116 Radishchev St., Kursk, 305004, Russian Federation,

Cand. Sci. (Engineering), Associate Professor, e-mail: [vvzhilin61@gmail.com](mailto:vvzhilin61@gmail.com), [https://elibrary.ru/author\\_profile.asp?id=:321882](https://elibrary.ru/author_profile.asp?id=:321882)

For classification of difficult structured data hybrid decisive modules with virtual streams are offered. Virtual streams reflect the hidden system communications between observed parameters of process or system. At the same time the vector of informative signs consists of two subvectors. First of which corresponds to real streams, and the second – to virtual streams. The generalized recurrent block diagram is developed for formation of a virtual stream. The scheme allows to form the hybrid vector of informative signs consisting of two subvectors. One of which consists of initial informative signs, and the second – of the informative signs, received on the basis of modeling of system communications between informative signs of the first subvector. For formation of the second subvector again created latent variables are used as system communications between initial space of informative signs. That allows to realize recurrent process of informative signs space formation. The offered method of nonlinear models formation for virtual streams is based on MGUA-modeling. For receiving models of real streams influence in the proposed method are used the virtual streams neural networks on nonlinear adalina. The method allows to form a subvector of latent variables of unlimited dimension. On the basis of the offered method the structure of hybrid decisive system with virtual streams, intended for classification of difficult structured data, is created. The structure allows to consider latent informative signs (virtual streams), defined on the basis of statistical and expert researches of communications between initial informative signs. In turn it gives the chance to aggregate accurate and indistinct decisive rules, providing the required quality of the decision making in the situations of diverse and badly formalizable structure of classes. Simulation showed the effectiveness of the proposed method.

**Keywords:** hybrid decision module, latent variable, GMDH model, neural network, fuzzy logic of decision-making, aggregators of fuzzy decision rules

**Введение.** Одним из важных направлений повышения качества управления лечебно-диагностическим процессом является эффективное прогнозирование и ранняя диагностика заболеваний, вызываемых профессиональной деятельностью. Характерной особенностью прогнозирования является то, что собираемая службами здравоохранения информация позволяет в основном анализировать уже имеющиеся заболевания, тогда как во многих случаях условия трудовой деятельности могут формировать предпосылки к возникновению и развитию той или иной патологии. Своевременное выявление людей, склонных к профессиональным заболеваниям и имеющих донозологические формы заболеваний, позволит формировать рациональные схемы организации лечебно-оздоровительных мероприятий, улучшающих качество медицинского обслуживания населения занятого в сельскохозяйственном и иных видах производств.

С математической точки зрения, сложность решения задач прогнозирования заболеваний определяется тем, что используемые информативные признаки носят неполный, нечеткий и разнотипный характер; имеют нечеткую структуру классов – это требует специальных подходов к синтезу соответствующих решающих правил. Раздельное использование таких мощных инструментариев, как нейронные сети и нечеткая логика не позволяет должным образом объединить экспертные знания и экспериментальные данные для решения рассматриваемой задачи. Поэтому возникает необходимость разработки методов объединения технологий нечеткой логики и нейронных сетей в гибридную систему для повышения качества прогнозирования профессиональных заболеваний.

Анализ существующих подходов к решению задач прогнозирования риска профессиональных заболеваний показал отсутствие серьезных системных исследований такого сложного явления, как «риск для жизни и здоровья человека, обусловленный производственной деятельностью хозяйствующих субъектов»; отсутствие единого понимания природы риска и менеджмента рисков. Такая ситуация препятствует внедрению методологии управления рисками в обеспечение безопасности производственной сферы в России. Поэтому необходимы дальнейшие исследования, направленные на разработку и совершенствование методов и моделей принятия решений для интеллектуальной системы поддержки прогнозирования профессиональных заболеваний [6, 16].

Проблема релевантности информативных признаков (ИП) является основной при построении классификаторов риска профессиональных заболеваний [2, 13]. Но не менее важной проблемой является решение о полноте пространства ИП, формируемого только из наблюдаемых переменных [16]. Известно, что в классифицируемом пространстве могут быть скрытые (ненаблюдаемые) или латентные закономерности. Их неучет приводит к построению ошибочной структуры классификатора и неправомерным выводам по анализу качества классификации построенной классифицирующей модели. Эта проблема усугубляется в условиях коротких выборок – для них могут быть также сделаны ошибочные выводы по показателям качества классификации выбранной модели принятия решений. Например, при ограничении объема выборки на качество восстановления влияния зависимости влияет не только качество аппроксимации в точках  $\{y_j\}$ , но еще и такие факторы, как сложность аппроксимирующей функции и размерности пространства  $N$ .

Такая особенность приводит к необходимости сосредоточить внимание на системных связях между наблюдаемыми переменными и латентными переменными. В свою очередь это позволит выбрать правильное соотношение между сложностью приближающей (аппроксимирующей) функции и объемом обучающей выборки (объемом имеющейся информации) [1, 8, 9, 14].

Исследования показали, что при построении решающих модулей по принятию решений для классификации в условиях неполной информации об объектах классификации и о структуре классов, целесообразно вводить дополнительные ИП, которые несут информацию о скрытых связях между исходными (наблюдаемыми) признаками. Основным требованием к этим связям является то, что они должны отличаться у исследуемых альтернативных классов. Скрытые связи определяются аппроксимирующей функцией, построенной по данным, извлеченным из обучающей выборки посредством формирования виртуальной переменной как зависимой переменной [7, 8, 9]. Если такая аппроксимирующая функция будет построена, то по ИП неизвестного образца может быть найден дополнительный признак (признаки) как функция этих входных данных, которые впоследствии используются в классификационных моделях в качестве независимых переменных.

**Методы и модели.** Среди немногих методов, в которых особое внимание уделяется поиску скрытых системных связей, выделяется метод группового учета аргументов (МГУА) [4, 5]. Метод МГУА позволяет получить самые разнообразные аппроксимирующие функции.

При поиске латентной переменной посредством МГУА необходимо решить задачу, состоящую в обнаружении и моделировании некоторой закономерности

$$y = F(X), \quad (1)$$

связывающей множество ИП  $X = (x_1, x_2, \dots, x_N)$ , используемых решающими модулями, с зависимой переменной  $y$ , принимающей в условии двухальтернативной классификации значения (0, 1).

Поиск закономерностей осуществляется в обучающей выборке, полученной для альтернативных классов, в которую входят  $M$  образцов из двух этих классов. Каждому образцу соответствует скалярная зависимая переменная  $y_j$  ( $j = \overline{1, M}$ ), принимающая значение ноль или единица. Модель зависимости (1), должна хорошо аппроксимировать ее не только в точках  $y_j$ , но и в любой точке пространства, определяемого координатами векторов  $X_i$ .

Рассмотрим процесс синтеза модели оптимальной сложности более подробно. Представим аппроксиматор (1) в общем виде

$$y = F(x_1, \dots, x_i, \dots, x_N), \quad (2)$$

где  $X_i$  –  $i$ -й информативный признак ( $i = 1, N$ ).

В качестве (2) можно использовать полином Колмогорова-Габор [4]:

$$y = \alpha_0 + \sum_{i=1}^N \alpha_i x_i + \sum_{i=1}^N \sum_{j=1}^N \alpha_{ij} x_i x_j + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \alpha_{ijk} x_i x_j x_k + \dots, \quad (3)$$

так как он позволяет добиться достаточно точной аппроксимации любой дифференцируемой функции  $F$ .

Сложная зависимость (3) заменяется множеством простых функций:

$$y_1 = f(x_1, x_2), \quad y_2 = f(x_1, x_3), \quad \dots, \quad y_s = f(x_{N-1}, x_N), \quad (4)$$

где  $s = C_N^2$ , причем вид функции  $f$  повсюду одинаков.

В качестве функции  $f$  обычно выбираются простые зависимости

$$y(x_i, x_k) = a_0 + a_1 x_i + a_2 x_k + a_3 x_i x_k \quad (5)$$

или

$$y(x_i, x_k) = a_0 + a_1 x_i + a_2 x_k + a_3 x_i x_k + a_4 x_i^2 + a_5 x_k^2, \quad (6)$$

связывающие только две переменные. Коэффициенты полиномов этих зависимостей можно определить по методу наименьших квадратов (МНК), имея соответственно 4 или 6 точек наблюдений в обучающей последовательности.

Среди моделей первого ряда выбираются  $S^*$  наилучших, показавших хорошие результаты на контрольной выборке. На следующем этапе (второй ряд алгоритма МГУА) полученные на обучающей выборке значения  $y_i$ , соответствующие отобраным моделям, рассматриваются в качестве аргументов нового ряда:

$$z_1 = f(y_1, y_2), \quad z_2 = f(y_1, y_3), \quad \dots, \quad z_{S^*} = f(y_{S^*-1}, y_{S^*}), \quad (7)$$

где  $S^* = C_S^2$ .

Здесь функции  $f$  остаются прежними и соответствуют соотношениям (6) или (7), но аргументами этих функций выступают переменные  $y_i$ . Коэффициенты новых моделей (7) также находятся по МНК на точках той же обучающей выборки. Новые модели проверяются на точках контрольной выборки, и среди них выбирается  $S^*$  наилучших, которые используются в качестве аргументов следующего, третьего ряда, и т.д. Сложность полиномов возрастает от ряда к ряду. Коэффициенты новых полиномов определяются по тем же точкам обучающей выборки, что не требует дополнительной информации, хотя сложность модели все время возрастает. При этом число определяемых коэффициентов значительно превосходит объем обучающей выборки.

Схема реализации алгоритма МГУА представлена на рисунке 1.

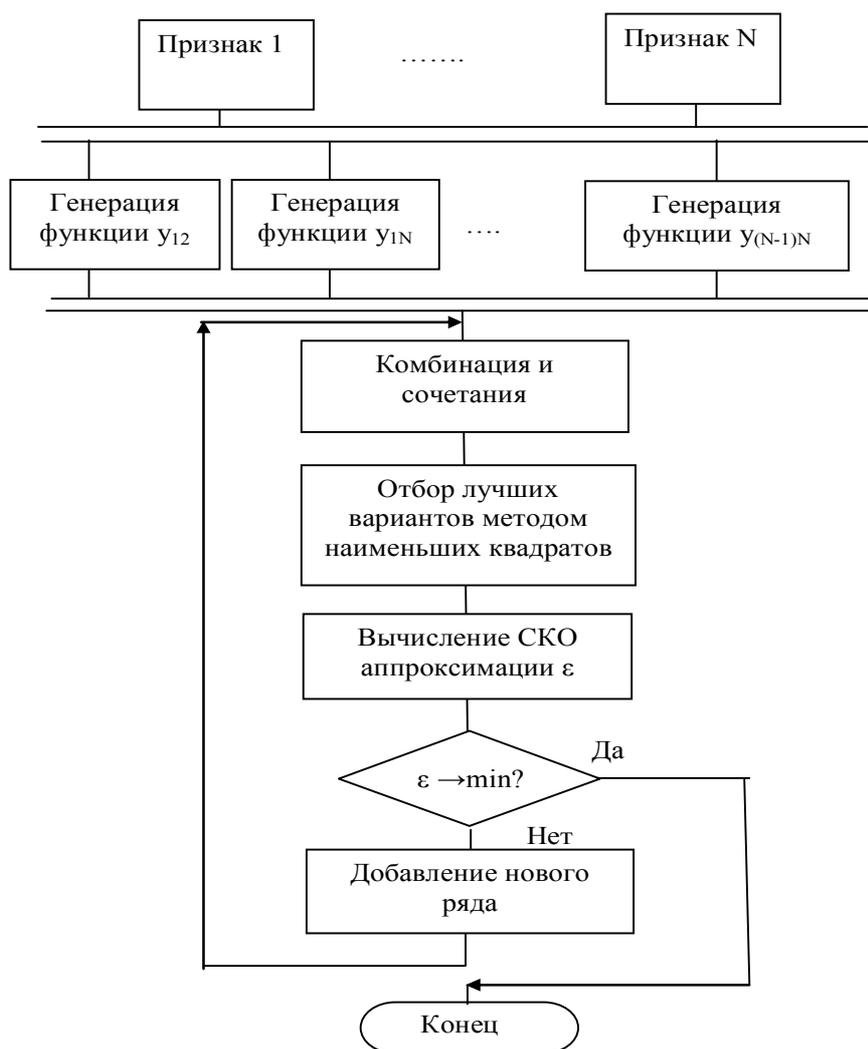


Рисунок 1 – Схема алгоритма реализации МГУА

В ней есть генераторы усложняющихся из ряда в ряд комбинаций и пороговые модули отбора лучших из них. Степенные полиномы заменяются несколькими рядами «частных» описаний. Например, (4) – ряд первой селекции и (7) – ряд второй селекции. Входные аргументы и промежуточные переменные сопрягаются попарно, и сложность комбинаций на каждом ряду обработки информации возрастает (как при массовой селекции), пока не будет получена единственная модель оптимальной сложности [1, 16].

При выполнении алгоритма по рисунку 1 из ряда в ряд селекция пропускает только некоторое количество самых адекватных переменных. Степень адекватности получаемых зависимостей оценивается по величине среднеквадратичной ошибки (СКО) аппроксимации  $\epsilon$  (средней для всех выбираемых в каждом поколении переменных или для одной самой точной переменной) на отдельной контрольной выборке.

Ряды селекции наращиваются до тех пор, пока точность аппроксимации повышается. Как только достигнут минимум ошибки, селекцию следует остановить [11].

Среди множества МГУА-моделей выбираем  $L$  наилучших, которые могут быть представлены в виде множества

$$\{f_{\ell}(X_k) = \tilde{y}\}, \tag{8}$$

где  $X_k \subseteq X$ ,  $X = (x_1, x_2, \dots, x_N)$  – множество ИП, используемых решающими модулями;  $\ell = \overline{1, L}$ .

Рассмотрим метод формирования множества (8). Метод должен синтезировать множество аппроксимирующих связей между элементами входного вектора  $X$  и позволить выбрать из этого множества  $L$  функциональных связей. Они формируют  $L$  дополнительных признаков, включение которых во входной вектор приводит к повышению качества классификации решающего модуля.

Включение в множество моделей очередной,  $(\ell + 1)$ -й, модели виртуального ИП осуществляется по рекуррентной схеме, представленной на рисунке 2.

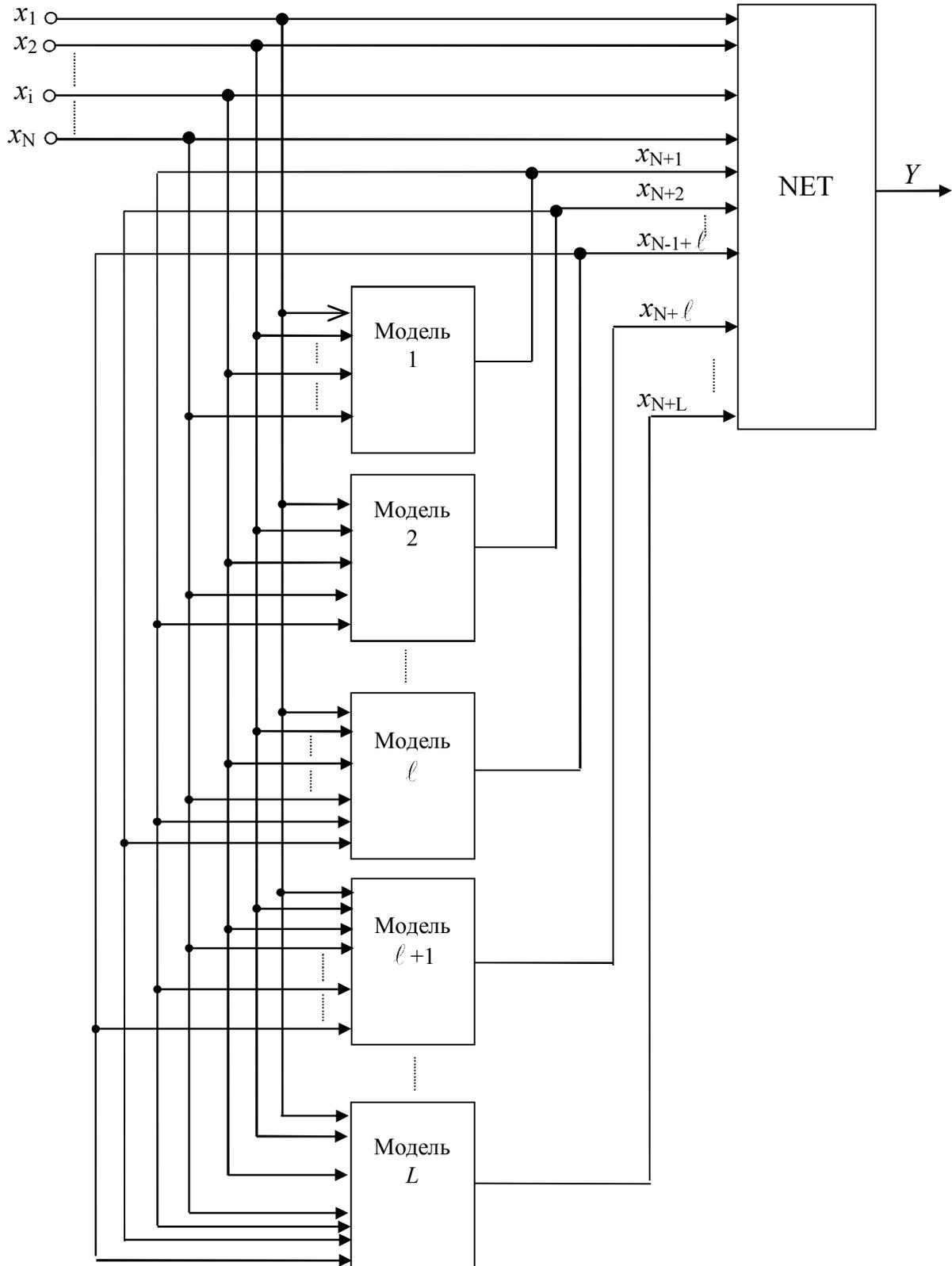


Рисунок 2 – Структурная схема классифицирующего модуля с дополнительными ИП, полученными на основе рекуррентных обратных связей

Эта схема позволяет оценить вклад в показатели качества принятия решений  $(\ell + 1)$ -го дополнительного ИП при наличии  $(\ell)$  дополнительных признаков.

**Метод формирования нелинейных моделей виртуальных потоков.** Для каждого виртуального потока посредством МГУА-моделирования получено свое подмножество моделей  $\{z_j\}_i \subseteq Z, \forall i = \overline{1, M}$ . Каждое подмножество моделей  $\{z_j\}_i, j = 1 \dots K_i$ , где  $K_i$  – число МГУ-моделей для  $i$ -го виртуального потока, полученных на основе МГУА-моделирования. Эти модели предполагается использовать для описания взаимного влияния известных ИП (реальных потоков) в системе простых комбинаций реальных и виртуальных потоков. Множество моделей виртуальных потоков  $\hat{Y}$  получается не посредством усложнения МГУА-моделей, а посредством МГУА-нейронной сети [11]. Отличие предлагаемой МГУА-нейронной сети от рассмотренной в [11] состоит в том, что каждый блок модели виртуального потока (второй слой модели) является МГУА-нейронной сетью. На ее входы поступают полученные путем МГУА-моделирования модели, включающие множество  $X$  реальных и виртуальных потоков. Структурная схема  $i$ -го блока модели виртуального потока представлена на рисунке 3.

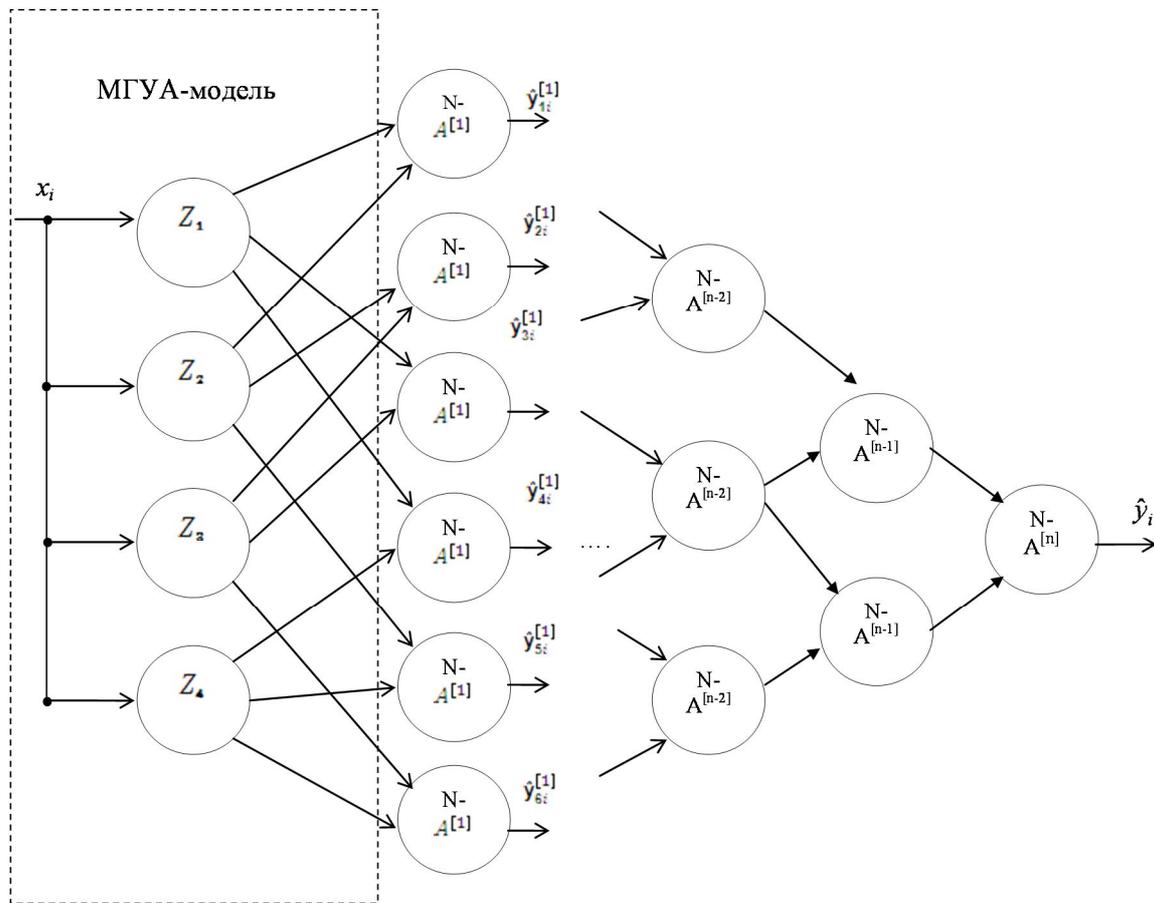


Рисунок 3 – Структурная схема  $i$ -й нелинейной модели первого слоя на основе МГУА – моделирования

В качестве примера в ней использованы четыре МГУА-модели виртуального потока  $x_i: z_1, z_2, z_3, z_4$  с нелинейными адалинами ( $N-A$ ).

МГУА-сеть имеет переменную структуру, которая может изменяться в процессе обучения. Каждый нейрон сети –  $N$ -адалина представляет собой адаптивный линейный ассоциатор с двумя входами  $z_g$  и  $z_h$  и нелинейным процессором, образованным тремя блоками умножения. Этот нейрон вычисляет квадратичную комбинацию входов вида

$$\hat{y}_{gh} = f_{gh}(z_g, z_h) = w_{gh_0} + w_{gh_1}z_g + w_{gh_2}z_g^2 + w_{gh_3}z_gz_h + w_{gh_4}z_h^2 + w_{gh_5}z_h = w_{gh}^T z_{gh}, \tag{9}$$

где  $w_{gh} = (w_{gh_0}, w_{gh_1}, w_{gh_2}, w_{gh_3}, w_{gh_4}, w_{gh_5})^T, z_{gh} = (1, z_g, z_g^2, z_gz_h, z_h^2, z_h)^T$ .

Процесс обучения МГУА-модели виртуального потока состоит в конфигурации нейронных сетей (рис. 4). Он начинается с первого скрытого слоя, независимой настройки синаптических весов каждой нелинейной адальны и наращивания количества слоев для достижения необходимой точности прогнозирования. Количество нейронов первого скрытого слоя сети определяется количеством  $K$  МГУА-моделей соответствующего эндогенного фактора и не превышает значение  $K(K-1)/2$  – количества сочетаний из  $K$  по 2. Обучение нейронной сети осуществляется посредством любого из известных алгоритмов обучения [13, 14].

Для количественного определения влияния дополнительного входа на качество классификации нейронной сети целесообразно два показателя (диагностическая чувствительность  $Se$  и диагностическая специфичность  $Sp$ ) выразить в виде одного. В качестве такого показателя часто используют диагностическую эффективность [3, 12]. Однако у этого показателя имеется существенный недостаток. Он связан с тем, что в результате простого объединения всех ошибок путем манипуляции с соотношениями положительных и отрицательных объектов в выборке этот показатель можно сделать достаточно большим. При этом классифицирующая система может быть полностью неработоспособна, например, иметь показатель диагностической чувствительности или диагностической специфичности, равный нулю. Поэтому введем интегральный показатель качества классификации  $S$ , способ вычисления которого поясняет рисунок 4.

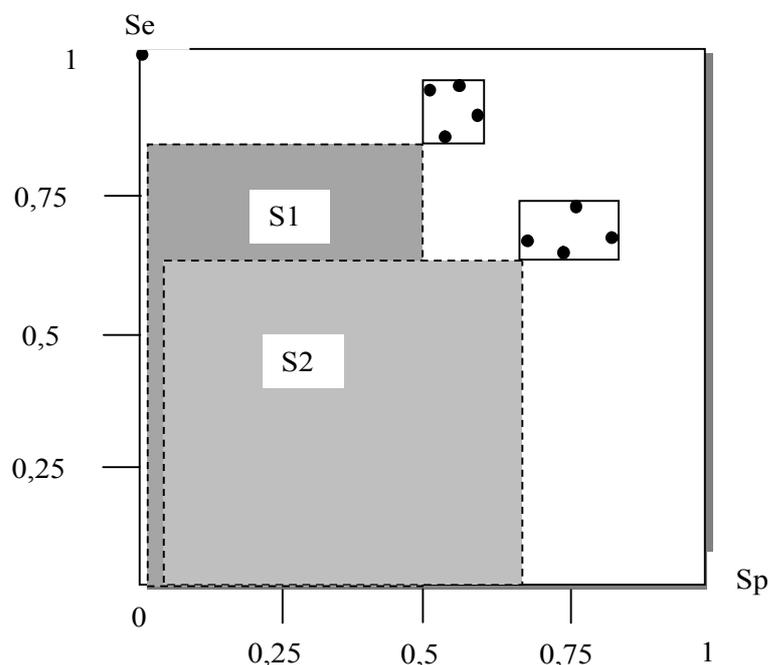


Рисунок 4 – Определение интегрального показателя качества классификации

Так же, как и в *ROC*-анализе, интегральный показатель качества классификации  $S_i$  ищем в единичном квадрате, определяющем координаты  $Sp$  и  $Se$ . Интегральный показатель  $S_i$  определяем по обучающим выборкам, компоненты векторов ИП которых принадлежат одной и той же генеральной совокупности.

На рисунке 4 показатели качества классификации определяются кружками черного цвета. Кружки, вписанные в прямоугольник, – это результаты классификации по выборкам из одной и той же генеральной совокупности. Для того чтобы не загромождать рисунок, в прямоугольнике показаны только четыре кружка. Они определяют координаты и размер прямоугольника (лежат на нижней, верхней, правой и левой границах).

Показатели  $S$  определяют площади заштрихованных прямоугольников, координаты которых задают левые нижние углы незаштрихованных прямоугольников, описывающих черные кружки. Для  $i$ -й выборки или группы выборок из одной и той же генеральной совокупности интегральный показатель качества классификации определяется по формуле

$$S_i = \sqrt{Se_{i \min} \cdot Sp_{i \min}}, \quad (10)$$

где  $Se_{i \min}$  и  $Sp_{i \min}$  определяются координатами описывающих прямоугольников.

Оценим степени управляемости показателями качества классификации гибридного решающего модуля посредством виртуального потока. С этой целью рассмотрим графики зависимости показателя качества классификации (10) от величины дисперсии виртуального входа, представленные на рисунке 5. Управляемость величиной  $S$  посредством изменения дисперсии виртуального потока исследовалась при различных размерностях пространства ИП: 3, 5 и 7.

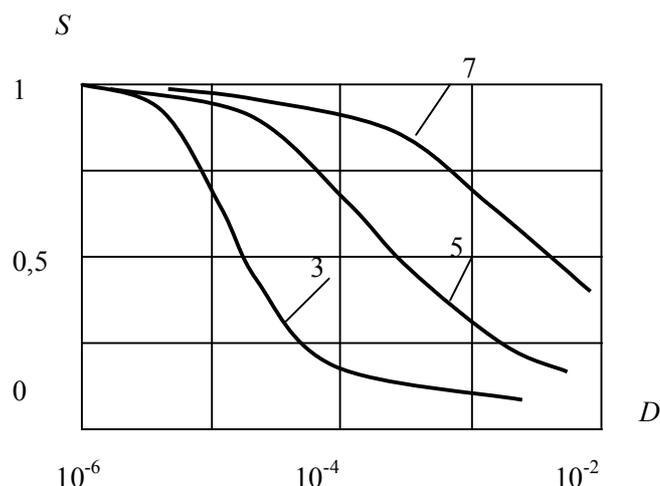


Рисунок 5 – Влияние дисперсии виртуального входа на качество классификации нейронной сети прямого распространения при размерности пространства ИП «три», «пять» и «семь»

Представленные на рисунке 5 зависимости показывают, что для широкого интервала значений размерности пространства ИП может быть подобрана дисперсия виртуального потока, которая позволяет управлять показателем качества классификации гибридного решающего модуля в широком диапазоне. При этом управляемость показателя качества классификации снижается с ростом пространства ИП. Отсюда следует, что при увеличении размерности пространства ИП необходимо увеличивать число виртуальных потоков.

**Модель гибридного классификатора с виртуальными потоками.** Анализ работы существующих методик прогнозирования показывает, что для повышения эффективности работы прогнозирующей системы необходимо объединение достоинств технологий нейронных сетей и систем нечеткого вывода [8, 12]. Для решения данной задачи разработана специализированная гибридная система. Она использует принципы обучения нейронных сетей на этапе дефuzziфикации, т.е. преобразования нечетких коэффициентов уверенности  $KU_i(\bar{x})$  в четкие номера классов  $\bar{y}_i$ .

Структура разработанной гибридной решающей системы для классификации набора объектов  $\{X_i\}$  представлена на рисунке 6. Система предназначена для прогнозирования профессиональных заболеваний. В данном случае рассматриваются бронхолегочные заболевания. В этой структуре три класса риска:  $C_1, C_2$  и  $C_3$ , двенадцать входных ИП  $P_1 \dots P_{12}$ , в которых выделено три сегмента  $G_1, G_2, G_3$ . В сегмент  $G_1$  вошли социально-экономические факторы риска, в  $G_2$  – микроклимат на рабочем месте, сегмент  $G_3$  характеризует индивидуальные факторы риска [6].

В качестве основы для разработки гибридной системы выбрана распространенная модель нечеткого решающего модуля, состоящая из блоков фуззификатора, агрегатора и дефuzziфикатора [12]. В отличие от основной структуры, в рассматриваемой системе на этапе фуззификации выполняется разбиение признакового пространства на группы для последующего анализа наборов сгруппированных признаков вместо совокупного анализа одновременно всех признаков исходного пространства. Агрегирование выполняется в два последовательных шага. На первом шаге выполняется построение структуры групповых агрегаторов, предназначенных для вычисления коэффициентов уверенности принадлежности объектов к заданному классу на основании каждой группы признаков.

На втором шаге для рассчитанных групповых коэффициентов уверенности строятся основные агрегаторы  $(A_1, A_2, A_3)$ , позволяющие непосредственно определить коэффициент уверенности принадлежности объекта к заданному классу. В качестве дефuzziфикатора гибридной системы используется нейронная сеть с выходами  $NET_1, NET_2, NET_3$ . Ее обучение производится на основе исходных обучающих данных, полученных из анализа образцов обучающей выборки больных бронхолегочными заболеваниями. В сформированное пространство информативных признаков, особенно в сегменты  $G_1$  и  $G_3$ , входят сильно скорректированные признаки, в то же время представленное признаковое пространство не исчерпывает всех факторов риска бронхолегочных заболеваний. Поэтому признаковое пространство было дополнено виртуальным потоком, который формируется нейронной сетью с выходом  $NET_4$ .

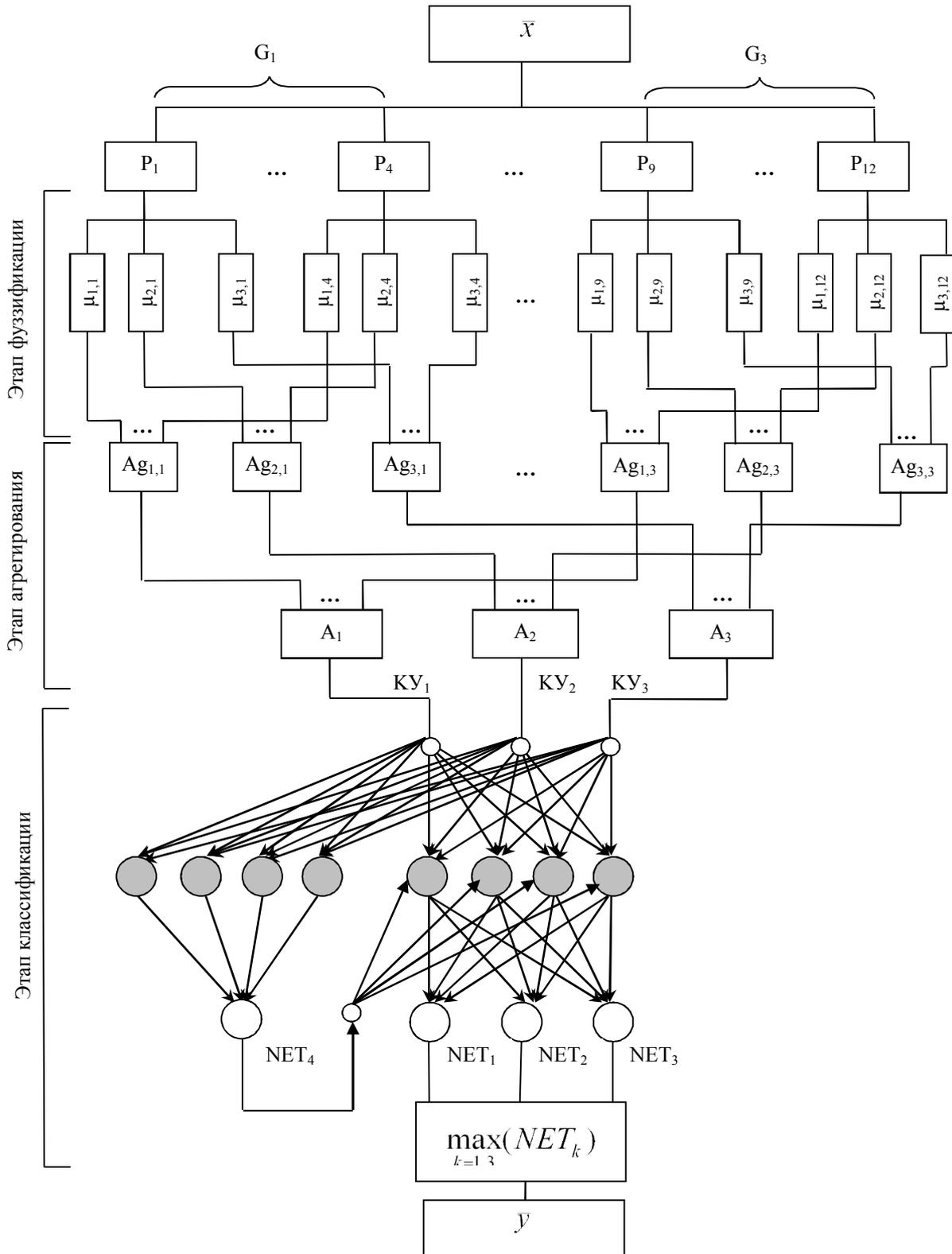


Рисунок 6 – Структура гибридной решающей системы с виртуальными потоками

Имитационное моделирование на основе искусственно сформированных выборок показало, что ввод («включение») дополнительного информативного признака (входа) в двухвходовой нейронной сети прямого распространения может повысить точность двухальтернативной классификации до 100 % и выше, при разнице первых моментов у третьего признака всего лишь на 10 %.

Был проведен анализ выборок экспериментальных данных, в которые входили 110 работников с установленным профессиональным риском бронхолегочных заболеваний и 150 относительно здоровых людей (не заболевших бронхолегочными заболеваниями в течение трех лет наблюдения). Результаты анализа показали, что на трехлетнем прогнозе диагностическая чувствительность (0,83) превосходит диагностическую специфичность (0,73) при невысоких значениях прогностической значимости ( $PZ^+$ ). Это соответствует преобладанию ошибки в отношении людей, оставшихся в состоянии «здоров».

Использование одного виртуального потока ( $NET_4$ ) позволило повысить диагностическую специфичность до 0,79 с оставшейся неизменной в пределах статистической погрешности вычислений диагностической чувствительностью.

**Выводы.** Для классификации плохоструктурированных данных предложены гибридные решающие модули с виртуальными потоками. Виртуальные потоки отражают скрытые системные связи между наблюдаемыми переменными. При этом вектор ИП состоит из двух подвекторов, первый из которых соответствует реальным потокам, а второй – виртуальным потокам.

Разработана обобщенная рекуррентная структурная схема формирования гибридного вектора ИП, который состоит из двух подвекторов. Один из них сформирован на основе исходных ИП. Второй – из ИП, полученных на основе моделирования системных связей между ИП первого подвектора. Указанная схема отличается использованием для формирования второго подвектора как системных связей между исходным пространством ИП, так и вновь сформированным признаковым пространством. В свою очередь это позволяет выделить релевантные системные связи в исходном пространстве ИП и во вновь сформированном пространстве ИП.

Разработан метод формирования нелинейных моделей виртуальных потоков, отличающийся использованием метода МГУА-моделирования для получения моделей влияния реальных потоков на виртуальные потоки, получаемые посредством нелинейных адалинов. Указанный метод позволяет формировать подвектор латентных переменных неограниченной размерности.

Предложена структура гибридной решающей системы с виртуальными потоками, предназначенная для классификации сложноструктурируемых данных. Указанная система отличается учетом в ее структуре латентных ИП (виртуальных потоков), определяемых на основе статистических и экспертных исследований связей между исходными информативными признаками. В свою очередь это позволяет агрегировать четкие и нечеткие решающие правила, обеспечивая требуемое качество принимаемых решений при разнородной и плохоформализуемой структуре классов.

#### Список литературы

1. Артеменко М. В. Формирование множества информативных показателей на основании аппроксимирующего полинома Колмогорова-Габора и максимального градиента функциональных различий / М. В. Артеменко, Н. М. Калугин, А. Н. Шуткин // Известия Юго-Западного государственного университета. Серия Управление, вычислительная техника, информатика. Медицинское приборостроение. – 2016. – № 1 (18). – С. 116–123.
2. Ахметвалеев А. М. Модель коллектива нейронных сетей и программный комплекс для определения функционального состояния человека / А. М. Ахметвалеев, А. С. Катасёв, М. А. Подольская // Прикаспийский журнал: управление и высокие технологии. – 2018. – № 1. – С. 69–85.
3. Жилин В. В. Способ моделирования нечетких моделей в пакете MATLAB для биомедицинских приложений / В. В. Жилин, С. А. Филлист, Халед Абдул Рахим и др. // Медицинская техника. – 2008. – № 2. – С. 15–17.
4. Петрова Т. В. Распределенные автономные интеллектуальные агенты для мониторинга и метаанализа эффективности управления живыми системами / Т. В. Петрова, А. А. Кузьмин, Д. Ю. Савинов и др. // Прикаспийский журнал: управление и высокие технологии. – 2017. – № 4 (40). – С. 61–73.
5. Позин А. О. МГУА-нейронные сети для прогнозирования состояния сложных систем с временными лагами / А. О. Позин, Е. А. Старцев, В. В. Уварова // Нейроинформатика, ее приложения и анализ данных: материалы XXIV Всероссийского семинара. – Красноярск: Институт вычислительного моделирования СО РАН, 2016. – С. 50–55.
6. Позин А. О. Метод синтеза трехуровневой нечеткой сети прогнозирования профессиональных заболеваний работников животноводческого комплекса / А. О. Позин // Энергосбережение и эффективность в технических системах: мат-лы IV Междунар. науч.-техн. конф. студентов, молодых ученых и специалистов. – Тамбов, 2017. – С. 51–54.
7. Позин А. О. Синтез дополнительного информативного признака для нейронной сети прямого распространения / А. О. Позин, С. А. Филлист, А. Н. Шуткин и др. // Новые решения в области упрочняющих технологий: взгляд молодых специалистов: сб. науч. ст. Междунар. науч.-практ. конф. – Курск: ЗАО «Университетская книга», 2016. – С. 212–217.
8. Позин А. О. Тестирующие гибридные системы с дополнительным пространством ИП / А. О. Позин, С. А. Филлист, А. Н. Шуткин // Современные информационные технологии в управлении качеством: сб. ст. V Междунар. науч.-приклад. конф. – Пенза: Приволжский Дом знаний, 2016. – С. 46–50.

9. Сидорова М. А. Алгоритм оценки качества работы нейронных сетей при решении задач классификации в медицинской диагностике человека / М. А. Сидорова, Н. А. Сержантова // Прикаспийский журнал: управление и высокие технологии. – 2017. – № 4. – С. 9–19.
10. Старцев Е. А. Двумерные нейросетевые структуры мета-анализа медико-экологических данных / Е. А. Старцев, О. В. Шаталова, В. В. Уварова // Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века : мат-лы Всерос. науч.-практ. конф. – Пермь : Перм. гос. нац. исслед. ун-т, 2016. – С. 161–164.
11. Томакова Р. А. Теоретико-множественный подход и теория графов в обработке сложноструктурируемых изображений : моногр. / Р. А. Томакова, О. В. Шаталова, М. В. Томаков. – Курск : Юго-Западный гос. ун-т, 2012. – 119 с.
12. Филист С. А. Гибридная нейронная сеть с макрослоями для медицинских приложений / С. А. Филист, О. В. Шаталова, М. А. Ефремов // Нейрокомпьютеры. Разработка и применение. – 2014. – № 6. – С. 35–39.
13. Филист С. А. Использование гибридных нейросетевых моделей для многоагентных систем классификации в гетерогенном пространстве ИП / С. А. Филист, А. Г. Курочкин, В. В. Жилин и др. // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 3 (31). – С. 85–95.
14. Филист С. А. Метод классификации сложных объектов на основе анализа структурных функций медленных волн / С. А. Филист, И. И. Волков, С. Г. Емельянов // Биомедицинская радиоэлектроника. – 2012. – № 4. – С. 6–11.
15. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – 2-е изд., испр. – Москва : ООО «И.Д. Вильямс», 2006. – 1104 с.
16. Шуткин А. Н. Адаптируемые к структуре данных классификаторы в задачах прогнозирования профессиональных заболеваний / А. Н. Шуткин, А. О. Позин, С. А. Филист // Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века : сб. ст. Второй Всерос. науч.-практ. конф. – Пермь : Перм. гос. нац. исслед. ун-т, 2017. – С. 161–164.

#### References

1. Artemenko M. V., Kalugin N. M., Shutkin A. N. Formirovanie mnozhestva informativnykh pokazateley na osnovanii approksimiruyushchego polinoma Kolmogorova-Gabora i maksimalnogo gradienta funktsionalnykh razlichiy [Formation of a set of informative indicators on the basis of the approximating Kolmogorov-Gabor polynomial and the maximum gradient of functional differences]. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Upravlenie, vychislitel'naya tekhnika, informatika. Meditsinskoe priborostroenie* [Proceedings of the Southwest State University. Series: Control, computer engineering, computer science. Medical devices], 2016, no. 1 (18), pp. 116–123.
2. Алхметвалеев А. М., Катасюв А. С., Подольская М. А. Model kolektiva neyronnykh setey i programmy kompleks dlya opredeleniya funktsionalnogo sostoyaniya cheloveka [Model of neural network collective and software package to determine the functional state of a person]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2018, no. 1, pp. 69–85.
3. Zhilin V. V., Filist S. A., Haled Abdul Rahim et al. Sposob modelirovaniya nechetkikh modeley v pakete MATLAB dlya biomeditsinskiykh prilozheniy [Method of modeling fuzzy models in MATLAB for biomedical applications]. *Meditsinskaya tekhnika* [Medical Technics], 2008, no. 2, pp. 15–17.
4. Petrova T. V., Kuzmin A. A., Savinov D. Yu. et al. Raspredeennye avtonomnye intellektualnye agenty dlya monitoringa i metaanaliza effektivnosti upravleniya zhivymi sistemami [Distributed autonomous intelligent agents for monitoring and metaanalysis of live systems management efficiency]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2017, no. 4 (40), pp. 61–73.
5. Pozin A. O., Startsev E. A., Uvarova V. V. MGUA-neyronnye seti dlya prognozirovaniya sostoyaniya slozhnykh sistem s vremennymi lagami [MGAA-neural networks for predicting the state of complex systems with time lags]. *Neyroinformatika, ee prilozheniya i analiz dannykh : materialy XXIV Vserossiyskogo seminar* [Neuroinformatics and its applications : Proceedings of the XXIV All-Russian Seminar]. Krasnoyarsk, Institute of Computational Modeling SB RAS, 2016, pp. 50–55.
6. Pozin A. O. Metod sinteza trekhurovnevoy nechetkoy seti prognozirovaniya professional'nykh zabolevaniy rabotnikov zhivotnovodcheskogo kompleksa [A method of synthesis three-level fuzzy network, forecasting of occupational diseases of cattle-breeding complex]. *Energoberezhenie i effektivnost v tekhnicheskikh sistemakh : materialy IV Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii studentov, molodykh uchenykh i spetsialistov* [Energy saving and efficiency in technical systems : Proceedings of the IV International Scientific and Technical Conference of Students, Young Scientists and Specialists]. Tambov, 2017, pp. 51–54.
7. Pozin A. O., Filist S. A., Shutkin A. N. et al. Sintez dopolnitelnogo informativnogo priznaka dlya neyronnoy seti pryamogo rasprostraneniya [Synthesis of additional informative feature for direct propagation neural network]. *Novye resheniya v oblasti uprochnyayushchikh tekhnologiy: vzglyad molodykh spetsialistov : sbornik nauchnykh statey Mezhdunarodnoy nauchno-prakticheskoy konferentsii* [New solutions in the field of strengthening technologies: the view of young professionals : Proceedings of the International Scientific-Practical Conference]. Kursk, JSC «Universitetskaya kniga» Publ., 2016, pp. 212–217.
8. Pozin A. O., Filist S. A., Shutkin A. N. Testiruyushchie gibridnye sistemy s dopolnitelnym prostranstvom informativnykh priznakov [Testing hybrid systems with additional space of informative features]. *Sovremennye informatsionnye tekhnologii v upravlenii kachestvom : sbornik statey V Mezhdunarodnoy nauchno-prikladnoy konferentsii* [Modern information technologies in quality management : Proceedings of the V International Scientific and Applied Conference]. Penza, Privolzhskiy Dom znaniy Publ., 2016, pp. 46–50.
9. Sidorova M. A., Serzhantova N. A. Algoritm otsenki kachestva raboty neyronnykh setey pri reshenii zadach klassifikatsii v meditsinskoy diagnostike cheloveka [The algorithm for assessing the quality of neural networks in solving classification problems in human medical diagnosis]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian journal: management and high technologies], 2017, no. 4, pp. 9–19.

10. Startsev E. A., Shatalova O. V., Uvarova V. V. Dvumernye neyrosetevye struktury metaanaliza mediko-ekologicheskikh dannykh [Two-dimensional neural networks structures meta-analysis physician-ecological data]. *Iskusstvennyy intellekt v reshenii aktualnykh sotsialnykh i ekonomicheskikh problem XXI veka : sbornik statey po materialam Vserossiyskoy nauchno-prakticheskoy konferentsii* [Artificial intelligence in the solution of urgent social and economic problems of the XXI century : proceedings of the All-Russian Scientific Practical Conference]. Perm, Perm State National Research University Publ., 2016, pp. 161–164.
11. Tomakova R. A., Shatalova O. V., Tomakov M. V. *Teoretiko-mnozhestvennyy podkhod i teoriya grafov v obrabotke slozhnostrukturiruemyykh izobrazheniy : monografiya* [A set-theoretic approach and graph theory in the treatment of difficult structured images : monograph]. Kursk, Publ. Southwest State University, 2012. 119 p.
12. Filist S. A., Shatalova O. V., Efremov M. A. Gibridnaya neyronnaya set s makrosloyami dlya meditsinskikh prilozheniy [Hybrid neural network with macro-layers for medical applications]. *Neyrokomp'yutery. Razrabotka i primeneniye* [Neurocomputers. Development and Application], 2014, no. 6, pp. 35–39.
13. Filist S. A., Kurochkin A. G., Zhilin V. V. et al. Ispolzovanie gibridnykh neyrosetevykh modeley dlya mnogooagentnykh sistem klassifikatsii v geterogenom prostranstve informativnykh priznakov [Use of hybrid neural network models for multi-agent classification systems in heterogeneous space of informative features]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2015, no. 3 (31), pp. 85–95.
14. Filist S. A., Volkov I. I., Emelyanov S. G. Metod klassifikatsii slozhnykh obektov na osnove analiza strukturnykh funktsiy medlennykh voln [Method of classification of complex objects based on the analysis of structural functions of slow waves]. *Biomeditsinskaya radioelektronika* [Biomedical Radioelectronics], 2012, no. 4, pp. 6–11.
15. Khaykin S. *Neyronnye seti: polnyy kurs* [Neural networks: a complete course]. 2nd ed., rev. Moscow, Publ. Ltd Co. «I.D. Vilyams» Publ., 2006. 1104 p.
16. Shutkin A. N., Pozin A. O., Filist S. A. Adaptiruemye k strukture dannykh klassifikatory v zadachakh prognozirovaniya professionalnykh zabolevaniy [Adapted to structure given qualifiers in problem of the forecasting of the professional diseases]. *Iskusstvennyy intellekt v reshenii aktualnykh sotsialnykh i ekonomicheskikh problem XXI veka : sbornik statey Vtoroy Vserossiyskoy nauchno-prakticheskoy konferentsii* [Artificial intelligence in the solution of urgent social and economic problems of the XXI century : Proceedings of the All-Russian Scientific Practical Conference]. Perm, Perm State National Research University Publ., 2017, pp. 161–164.

УДК 004.89

## МЕТОДЫ И МОДЕЛИ ДЛЯ АВТОМАТИЗИРОВАННОЙ ДИАГНОСТИКИ БОЛЕЗНИ ПАРКИНСОНА

*Статья поступила в редакцию 09.10.2018, в окончательном варианте – 10.11.2018.*

**Жвалецкий Олег Валерьевич**, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), 199178, Российская Федерация, г. Санкт-Петербург, 14-я линия В.О., 39, научный сотрудник, e-mail: ozh@spiiras.ru

Рассмотрена проблема автоматизации диагностики болезни Паркинсона, основанной на математическом анализе физиологических сигналов. Целью автоматизации является создание инструмента для объективной, точной и, по возможности, ранней диагностики болезни Паркинсона. Показано, что для решения каждой из этих задач требуется создание/использование различного аппаратного и математического обеспечения. Объективная диагностика болезни Паркинсона осуществима при помощи простого приложения, которое может функционировать на любом смартфоне, оснащённом стандартным набором датчиков. Точная диагностика болезни Паркинсона требует специализированного измерительного оборудования, гарантирующего высокое качество результатов измерений, специальным образом организованного измерительного эксперимента, обеспечивающего получение достаточно полной и достоверной информации о процессах регуляции движений, а также специального математического обеспечения, реализующего методы машинного обучения и методы математического моделирования. Ранняя диагностика болезни Паркинсона оказывается возможной только при комплексировании различных видов измерений, в т.ч. и с привлечением данных, соответствующих немотротным проявлениям болезни Паркинсона. Обоснована целесообразность построения автоматизированной системы диагностики болезни Паркинсона на базе математического анализа тензотримограмм, регистрируемых по специальной методике, предложенной С.П. Романовым для объективной оценки состояния системы построения движений по Н.А. Бернштейну. Сформулирована основная цель исследования – построение системы распознавания, основанной на математическом анализе тензотримограмм. Предложена многоуровневая архитектура такой системы распознавания и показано место в этой системе методов машинного обучения и методов математического моделирования.

**Ключевые слова:** болезнь Паркинсона, диагностика заболеваний, автоматизация, функциональное состояние, тензотримограмма, анализ временных рядов, машинное обучение, математическое моделирование, система распознавания, концептуальная модель