

УДК 004.82

doi: 10.15622/rcai.2025.039

ИНТЕРПРЕТИРУЕМЫЙ МНОГОСЛОЙНЫЙ КЛАССИФИКАТОР НА ОСНОВЕ НЕЧЕТКО-НЕЙРОСЕТЕВОЙ КОГНИТИВНОЙ МОДЕЛИ В ЗАДАЧАХ ТЕХНИЧЕСКОЙ ДИАГНОСТИКИ

Я.С. Нападайло (*napadailo@skynapp.com*)

В.А. Тайлаков (*vic7519@mail.ru*)

Г.В. Рыбина (*gvrybina@yandex.ru*)

Национальный исследовательский ядерный университет МИФИ,
Москва

В работе рассматривается интерпретируемый многослойный классификатор на основе нечетко-нейросетевой когнитивной модели для решения задач технической диагностики, в частности диагностики неисправностей ветрогенератора, а также оценивается эффективность и точность обнаружения неисправностей в сравнении с классификатором на базе рекуррентной нейронной сети. Сравниваются различные подходы к интерпретации вывода моделей типа “черный ящик”, применяемых в задачах технической диагностики, и дополненных нечеткой когнитивной моделью для объяснения результатов их работы, с подходом к интерпретации результатов диагностики неисправностей на основе предлагаемой авторами модели.

Ключевые слова: техническая диагностика, интерпретируемость, нечеткая когнитивная модель, интеллектуальные системы поддержки принятия решений.

Введение

В задачах технической диагностики и связанных с ними задачах управления объектами технической инфраструктуры в настоящее время широко применяются гибридные интеллектуальные системы, в частности интеллектуальные системы поддержки принятия решений для динамических областей, особенно в условиях неопределенности, нелинейности, зашумленности [Рыбина, 2014]. Непосредственный интерес для исследователя представляют методы “черного ящика”, возникшие как следствие

второй и третьей волн искусственного интеллекта [Забежайло и др., 2022]: глубокие нейронные сети, рекуррентные нейронные сети, нейро-нечеткие модели и др. На основе данных методов разрабатываются различные компоненты интеллектуальных систем, направленные на решение специфических задач в рамках гибридного подхода, обеспечивающего синергетический эффект в результате объединения двух и более диагностических парадигм.

Однако, с ростом популярности применения моделей “черного ящика” к задачам технической диагностики возникает беспокойство относительно причин, побудивших модель принять то или иное решение [Аверкин, 2023]. Эта проблема особенно актуальна для критически важных объектов технической инфраструктуры: оператор мониторинга должен получать как можно больше контекстной информации, касающейся возможных сбоев или аварийных ситуаций, и иметь основание доверять этим данным. Интерпретируемость как основа доверия человека интеллектуальной системе поддержки принятия решений является одним из ключевых принципов третьей волны искусственного интеллекта. Довольно перспективным и интересным направлением в области разработки интерпретируемых моделей становится исследование алгоритмов и методов нечеткой логики, в частности, нейро-нечетких моделей [Ефремова и др., 2017], а также обучаемых нечетких когнитивных моделей. Можно выделить три основные архитектуры построения интерпретируемых моделей на основе нейро-нечетких систем по типу взаимодействия нечеткой системы и модели “черного ящика”: кооперативные, параллельные и гибридные нейро-нечеткие модели [Аверкин, 2024]. В литературе также описаны интерпретируемые полиморфные гибридные модели для решения задач технической диагностики: например, в работе [Mansouri et al., 2023] поведение рекуррентной нейронной сети (РНС) имитируется с помощью нечеткой когнитивной карты (НКК), обученной на тех же данных, что и РНС.

В данной работе рассматривается интерпретируемый многослойный классификатор на основе нечетко-нейросетевой когнитивной модели для решения задач технической диагностики, в частности диагностики неисправностей ветрогенератора, а также оценивается эффективность и точность обнаружения неисправностей по сравнению с классификатором на базе рекуррентной нейронной сети. Сравниваются различные подходы к интерпретации вывода моделей типа “черный ящик”, применяемых в задачах технической диагностики, и дополненных нечеткой когнитивной моделью для объяснения результатов их работы, с подходом к интерпретации результатов диагностики неисправностей на основе предлагаемой авторами модели.

1. Нечетко-нейросетевая когнитивная модель

Нечеткие когнитивные модели (НКМ) [Захарова и др., 2020], в частности НКК, в качестве инструмента поддержки принятия решений находят широкое применение в экономике – для анализа сложных систем в условиях риска и неопределенности [Заграновская, 2018], на производстве – для анализа причин и последствий отказов технических систем в динамических средах [Wang et al., 2022], а также для мониторинга и управления техногенными рисками на предприятиях [Борисов, 2020a], в энергетике и робототехнике – для решения задач диагностики и управления сложными техническими объектами в динамических недетерминированных средах [Karatzinis et al., 2025].

Одной из перспективных разновидностей НКМ с точки зрения ее применения в задачах технической диагностики и управления, является обучаемая модель FCN [Karatzinis et al., 2023]. Отличительной особенностью модели является ее способность функционировать во взаимосвязи с моделируемой динамической системой, что позволяет “дообучать” параметры модели в процессе ее эксплуатации. Согласно классификации, основанной на систематизации нечетких когнитивных моделей [Борисов, 2020b] модель FCN следует отнести к гибридным нечетким моделям “с функциональным замещением”:

- в качестве основной модели берется НКК, которую согласно предложенной в [Кулинич, 2011] классификации когнитивных карт можно отнести к детерминированным качественным когнитивным картам, основанным на правилах, по двум основным классифицирующим признакам;
- в качестве дополнительных компонентов выступают линейный и билинейный адаптивные алгоритмы обучения весов и параметров активационных функций.

Таким образом модель FCN является гибридной нечетко-нейросетевой моделью с параметрической оптимизацией весов и активационных функций на основе алгоритмов обучения с использованием обучающей выборки.

Зададим модель *FCN* в общем виде:

(1.1)

где A – множество концептов; C – вектор коэффициентов угла наклона функции активации; W – матрица весов; $Z(a)$, $Z(w)$ и $Z(c)$ – шкалы значений концептов, силы влияния концептов и значений углов наклона сигмоидальной функции активации; R – база нечетких правил, которая хранит параметры стационарных состояний динамической системы, L – оператор обучения модели (LPM – линейный, BPM – билинейный). Шкалы значений концептов и силы влияния представлены в виде отображений:

$$, \quad (1.2)$$

$$, \quad (1.3)$$

$$, \quad (1.4)$$

где μ_A , μ_W и μ_C – функции принадлежности, отображающие значения концептов, весов и углов наклона активационных функций в степени принадлежности соответствующим нечетким подмножествам шкалы X . Множество нечетких правил R , устанавливающее соответствие между значениям концептов A , весов W и углов наклона активационных функций C и некоторым стационарным состоянием S , задается в виде:

$$(1.5)$$

Алгоритм построения модели FCN включает два этапа: структурную и параметрическую идентификацию. На первом этапе эксперт задает структуру НКК: множество концептов A и наличие причинно-следственных связей W между ними. Множество концептов задается в виде:

$$, \quad (1.6)$$

где U – множество входных концептов (управляемых и внешних), а V – множество всех остальных концептов (целевых и наблюдаемых). Входной концепт не имеет входящей причинной связи от других концептов и не меняет свое значение в процессе сходимости НКК к точке равновесия. На втором этапе выполняется параметризация НКК посредством обучения её весов на обучающей выборке и последующего за ним формирования базы нечетких правил. Для обучения весов и углов наклона активационных функций в модели FCN применяются линейный и билинейный адаптивные алгоритмы. Линейный адаптивный алгоритм LPM обучения модели FCN описывается системой уравнений [Boutalis et al., 2014]:

$$\frac{dw_{ij}}{dt} = -w_{ij} + \eta \delta_i x_j, \quad (1.7)$$

$$, \quad (1.8)$$

где δ_i – ошибка оценки строки i матрицы весов W на шаге k , η – обратное значение функции активации от заданной точки равновесия системы, w_{ij} – строка матрицы весов W на шаге k , x_j – вектор значений концептов в заданной точке равновесия системы, c – гиперпараметр, α – скорость обучения, $c - \alpha$ – обновленная строка матрицы W на шаге k . Для обеспечения сходимости и устойчивости НКК применяются методы ортогональной проекции на выпуклые множества S и R , заданные формулами (1.9) и (1.10).

, (1.9)

где w_{ij} – значение веса для пары концептов C_i, C_j , M – ограничение по модулю веса w_{ij} .

, (1.10)

где \mathbf{Z} – вектор столбец, элементами которого являются строки матрицы весов \mathbf{W} , $\|\mathbf{Z}\|$ – евклидова норма вектор-столбца, а M – ограничение по норме $\|\mathbf{Z}\|$.

Для уменьшения размера НКК (количества концептов) применяется билинейный адаптивный алгоритм обучения *BPM*, учитывающий вектор коэффициентов \mathbf{C} углов наклона функции активации (сигмоиды).

Расширим модель *FCN* (1.1), добавив в ее описание интерпретируемый компонент T :

(1.11)

где T – матрица взаимных влияний концептов, размерность которой равна размерности матрицы весов, $Z(t)$ – шкала значений взаимных влияний концептов. Пусть m – количество всех путей между двумя концептами C_i и C_j . Тогда косвенным эффектом влияния для пары концептов C_i, C_j по пути l , проходящему через промежуточные концепты $C_{i_1}, C_{i_2}, \dots, C_{i_{k-1}}$ будем называть минимальное значение веса среди всех смежных пар концептов $C_i, C_{i_1}, C_{i_1}, C_{i_2}, \dots, C_{i_{k-1}}, C_j$ на этом пути. Общим эффектом влияния для пары концептов C_i, C_j будем считать максимальное из значений косвенных эффектов влияния E_{ij}^l , по всем m возможным путям от C_i к C_j . Общие эффекты влияния для каждой пары концептов C_i, C_j вычисляются с помощью аппарата нечеткой причинной алгебры [Kosko, 1986]:

(1.12)

(1.13)

2. Интерпретируемый многослойный классификатор

Под диагнозом будем понимать класс отклонений параметров динамической системы от их номинальных значений. Тогда задачу технической диагностики можно свести к задаче классификации, а задачу выявления причин, приведших к возникновению этих отклонений – к задаче интерпретации результатов классификации. Для этого построим классификатор, состоящий двух и более слоев, где каждый отдельный его слой представлен одной моделью *FCN*. Для большей ясности обратимся вновь к работе [Борисов, 2020b] и поясним, что из себя представляет этот классификатор согласно систематизации гибридных нечетких моделей. Поскольку мо-

дель *FCN*, представляющая отдельный слой, является гибридной нечетко-нейросетевой моделью с функциональным замещением, то предлагаемый многослойный классификатор будет ни что иное, как композиционная гибридная нечетко-нейросетевая модель для поддержки принятия решений в системах диагностики.

В данной работе ограничимся следующим набором подзадач для решения комплексной задачи технической диагностики: а) обнаружение дефекта (отклонения); б) локализация дефекта (отклонения); в) выявление причины, приведшей к неисправности и/или обоснование принятого решения. Для решения каждой из подзадач будем использовать отдельную модель *FCN*, формирующую отдельный слой классификатора. Методы, выбранные для решения каждой из подзадач, а также буквенные обозначения слоев приведены в табл. 1.

Таблица 1

№	Слой	Подзадача	Метод
#1		Обнаружение неисправности	Бинарная классификация
#2		Локализация неисправности	Многоклассовая классификация
#3		Интерпретация результатов	Статический анализ

Обозначим модель классификатора как *FCNXAI* и опишем структуру модели:

$$FCNXAI = \{FCN_1, FCN_2, FCN_3\}, \quad (2.1)$$

где слой FCN_1 – модель *FCN* обнаружения дефекта, которая соотносит текущее состояние объекта диагностирования к одному из режимов функционирования – штатному либо аварийному; слой FCN_2 – модель *FCN* локализации дефекта, D – множество дефектов (классов отклонений) объекта диагностирования, n – общее число дефектов (классов отклонений), S – множество признаков (симптомов) объекта диагностирования, m – общее число признаков; слой FCN_3 – модель *FCN* интерпретации результатов диагностирования.

Построение классификатора выполняется следующим образом. Для слоя FCN_1 сначала выполняется структурная идентификация модели *FCN*: эксперт задает множество концептов U в виде:

$$U = \{U_1, U_2, \dots, U_n\}, \quad (2.2)$$

где U – множество управляемых и внешних концептов; S – множество признаков (симптомов); M – целевой концепт, который задает режим функционирования системы (штатный или аварийный). Далее веса

обучаются с помощью алгоритма *LPM* на статистических данных. Затем эксперт выполняет структурную идентификацию слоя . Множество концептов задается в виде:

$$U = \{u_1, u_2, \dots, u_n\}, \quad (2.3)$$

где U – множество входных концептов; S – множество признаков (симптомов); D – множество дефектов (классов отклонений). После этого веса слоя обучаются на статистической выборке с помощью алгоритма *LPM*. Множество концептов слоя задается как:

$$R = \{r_1, r_2, \dots, r_m\}, \quad (2.4)$$

где R – множество причин, приведших к неисправности, D – множество дефектов (классов отклонений). Матрицы и слоя обучаются на статистической выборке с помощью алгоритма *LPM*. Наконец, модель *FCN* любого из слоев классификатора *FCNXAI* может быть оптимизирована (редуцирована) при необходимости с помощью алгоритма *BPM*.

Классификатор *FCNXAI* может применяться в задачах технической диагностики по следующему сценарию: на вход слоя подается вектор значений концептов , затем слой выполняет задачу обнаружения неисправности. Результат бинарной классификации M передается на слой . Если выявлено, что система работает в аварийном режиме – слой выполняет задачу локализации неисправности и определяет класс неисправности (или несколько классов) системы: для этого на вход слоя подается вектор значений концептов . Далее результат многоклассовой классификации D , а также вектор значений концептов подается на слой , который интерпретирует полученный на предыдущем шаге результат.

3. Анализ различных подходов к интерпретации в задачах технической диагностики

Примеры интерпретируемых моделей “черного ящика” в различных приложениях поддержки принятия решений подробно описаны в литературе. Так, в работе [Wang et al., 2021] представлена гибридная нечетко-нейросетевая модель *DFCM* на основе НКК для прогнозирования загруженности автомобильных дорог; веса модели представляют собой обученные глубокие нейронные сети (ГНС), моделирующие нелинейные зависимости динамической системы, а также включают рекуррентные нейронные сети (РНС), учитывающие влияние внешних возмущений. В [Mansouri et al., 2023] гибридная полиморфная нечеткая когнитивная модель *LFCM*, поддерживающая интерпретацию вывода РНС, обучается на тех же данных, что и модель РНС; модель *LFCM* применяется в задаче диагностики неисправностей ветрогенератора. Отметим, что в обоих случаях описанная модель представляет собой результат гибридизации ней-

ронной сети, моделирующей динамическое поведение системы, и НКК, описывающей причинно-следственные связи между отдельными компонентами системы (концептами).

В предложенном подходе нелинейные отношения между концептами каждого слоя классификатора моделируются с помощью множества нечетких правил R , которые описывают параметры стационарных состояний динамической системы; при этом алгоритмы LPM и BPM гарантируют сходимость модели FCN в процессе ее обучения, а также обеспечивают взаимодействие с моделируемой динамической системой в режиме реального времени. Встроенный механизм интерпретации результатов диагностики, в рамках системы поддержки принятия решений, позволяет выявить причину, приведшую к неисправности и предоставить оператору мониторинга обоснование принятого решения. Кроме того, модель обладает низкой вычислительной нагрузкой, что делает ее применимой в приложениях реального времени.

Рассмотрим применение предложенного классификатора на примере решения двух задач диагностики, описанных в подразделах 3.1 и 3.2.

3.1. Задача диагностики неисправностей подшипников

В качестве набора данных для проведения эксперимента был взят публичный датасет со значениями, снятыми с датчиков вибраций, собранными с 4-х промышленных подшипников на протяжении 4-х месяцев. Все подшипники имели одинаковые характеристики (класс 1, до 15 кВт), а в качестве порога вибрации для данного типа подшипников было выбрано среднее квадратичное значение виброскорости 0.5 мм/с согласно стандарта ГОСТ ИСО 10816-1-97. Три подшипника находились в исправном состоянии, четвертый подшипник - неоднократно выходил из строя.

Для задачи диагностики неисправностей подшипников был построен классификатор $FCNXAI$, обученный на статистической выборке из экспериментального набора данных с помощью алгоритма . После обучения модель $FCNXAI$ была проверена на тестовом наборе данных. Для проведения сравнительной оценки эффективности модели был выбран классификатор $LFCM$. Модель $LFCM$ демонстрирует немного более высокую точность на тестовой выборке; в то же самое время предлагаемая модель $FCNXAI$ обеспечивает существенно меньшую вычислительную сложность, а также высокую интерпретируемость (рис. 1). В отличие от модели $LFCM$, в которой элементы матриц весов W и T имеют фиксированные значения после обучения, модель $FCNXAI$ сохраняет нелинейные отношения в виде множества нечетких правил R , обеспечивая более точную и контекстно-зависимую интерпретацию в условиях выраженной нелинейности исходных данных.

Таблица сравнения:

	Train Time (s)	Accuracy	AUC	Inference Time (ms/sample)
RNN	84.7899	0.9763	0.9257	0.0228
CNN	107.3366	0.9755	0.9379	0.0425
LFCM	56.6601	0.9744	0.9426	0.0164
FCNXLAI	50.4289	0.9507	0.9002	0.0150

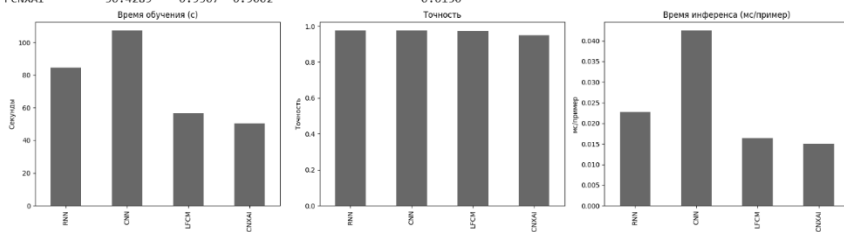


Рис. 1. Сравнение быстродействия моделей

3.2. Задача диагностики неисправностей обмотки генератора

Для решения задачи диагностики неисправностей обмотки генератора была построена тепловая модель *FCNXLAI*, учитывающая влияние внешних факторов, таких как скорость ветра, температура окружающей среды и влажность. Внешние факторы представлены множеством входных концептов U , тогда как все остальные концепты, включая значения температуры обмоток генератора, образуют множество V . Значения внешних факторов для обучающей и тестовой выборок были взяты из метеорологической базы данных gistemeo.kz для города Алматы (Казахстан), а значения температуры обмоток генератора получены с температурных датчиков, установленных внутри корпуса 3-х фазного генератора с прямым приводом мощностью 1 кВт и частотой вращения вала 300 об/мин.

Модель позволяет заранее предсказать возможный перегрев обмоток с учетом изменения внешних факторов окружающей среды и эксплуатационных режимов, а также помогает выяснить, какие именно внешние факторы вносят наибольший вклад в рост температуры.

При проведении статического (структурно-целевого) и динамического (сценарного) анализа было установлено, что наиболее критичным является сочетание высоких температур воздуха () и низкой скорости ветра (менее 3 м/с), при котором температура обмоток может достигать , превышая допустимые значения для некоторых стандартных классов изоляции (ГОСТ Р МЭК 60085–2011). На рис 2 приведён анализ влияния внешних факторов на температуру обмотки генератора в виде радиальной диаграммы, включающей восемь концептов (A1–A8), где A1 – температура обмотки, A2 – скорость ветра, A3 – влажность, A8 – температура воздуха, A4–A7 – наблюдаемые параметры. В другом сценарии анализ показал, что повышенная влажность воздуха (более 70%) увеличивает температуру обмоток на 5–10% за счет ухудшения теплоотдачи, связанного с изменением термических свойств воздуха.

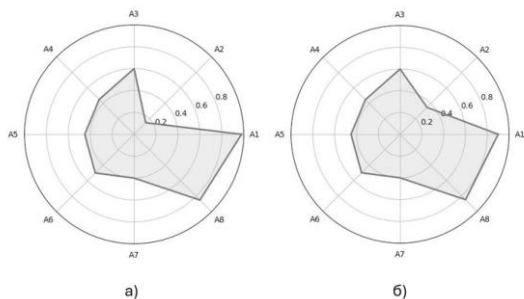


Рис. 2. Анализ влияния внешних факторов на температуру обмотки:
а) низкая скорость ветра, б) средняя скорость ветра

Заключение

В данной работе рассмотрен интерпретируемый многослойный классификатор *FCNXAI* на основе гибридной нечетко-нейросетевой модели для решения задач технической диагностики, в частности диагностики неисправностей ветрогенератора. Предлагаемый классификатор, с одной стороны, позволяет моделировать сложные нелинейные зависимости динамической системы, а также обеспечивает сходимость модели во время ее обучения, а с другой – предоставляет оператору мониторинга интерпретируемый вывод, обоснование принятого моделью решения. Эксперименты показали, что модель обладает высокой интерпретируемостью и более низкой вычислительной способностью по сравнению с другими гибридными нечетко-нейросетевыми моделями, такими как *LFCM*, что делает возможным ее применение в задачах реального времени.

Дальнейшие исследования будут направлены на изучение взаимодействия двух и более классификаторов, предназначенных для решения задачи диагностики отдельных компонентов сложной динамической системы. Кроме того, планируется уделить внимание исследованию методов оптимизации (упрощения структуры) модели с помощью алгоритма *BPM*.

Список литературы

- [Аверкин, 2023] Аверкин А.Н. Объяснимый искусственный интеллект как часть искусственного интеллекта третьего поколения // Речевые технологии. – 2023. – № 1. – С. 4-10.
- [Аверкин, 2024] Аверкин А.Н. Объяснительный искусственный интеллект в больших речевых моделях // Речевые технологии. – 2024. – № 1. – С. 3-13.
- [Борисов, 2020а] Борисов В.В. Нечёткие когнитивные модели как основа для исследования сложных систем и процессов // Речевые технологии. – 2020. – № 1-2. – С. 48-62.

- [Борисов, 2020b] Борисов В.В. Систематизация нечетких и гибридных нечетких моделей // Мягкие измерения и вычисления. – 2020. – Т. 29, № 4. – С. 98-120.
- [Ефремова и др., 2017] Ефремова Н.А., Аверкин А.Н., Ярушев С.А. Гибридные нечёткие когнитивные карты в задачах поддержки принятия решений и прогнозирования // Программные продукты, системы и алгоритмы. – 2017. – № 4. – С. 18-25.
- [Захарова и др., 2020] Захарова А.А., Подвесовский А.Г., Исаев Р.А. Нечеткие когнитивные модели в управлении слабоструктурированными социально-экономическими системами // Информационные и математические технологии в науке и управлении. – 2020. – № 4 (20). – С. 5-23.
- [Забейайло и др., 2022] Забейайло М.И., Борисов В.В. Об интерпретациях понятия “искусственный интеллект” // Речевые технологии. – 2022. – № 1. – С. 5-18.
- [Заграновская, 2018] Заграновская А.В. Системный анализ на основе нечётких когнитивных карт // Вестник Российского экономического университета имени Г.В. Плеханова. – 2018. – № 4. – С. 152-160.
- [Кулинич, 2011] Кулинич А.А. Классификация когнитивных карт и методы их анализа // 6-я международная научно-техническая конференция “Интегрированные модели и мягкие вычисления в искусственном интеллекте” (Коломна, 2011 г.). Труды конференций. Т. 1 – М.: Физматлит, 2011. – С. 124-135.
- [Рыбина, 2010] Рыбина Г.В. Основы построения интеллектуальных систем. – М: Финансы и статистика; ИНФРА-М, 2010. – 432 с.
- [Boutalis et al., 2014] Boutalis Y., Theodoridis D., Kottas T., Christodoulou M.A. Theory and Applications of the Neurofuzzy and Fuzzy Cognitive Network Models. System Identification and Adaptive Control. – Springer International Publishing, 2014. doi: 10.1007/978-3-319-06364-5.
- [Karatzinis et al., 2023] Karatzinis G.D., Apostolikas N.A., Boutalis Y.S., Papakostas G.A. Fuzzy Cognitive Networks in Diverse Applications Using Hybrid Representative Structures. International Journal of Fuzzy Systems. –2023. – Vol. 25. – P. 2534-2554. – doi: 10.1007/s40815-023-01564-4.
- [Kosko, 1986] Kosko B. Fuzzy Cognitive Maps // International Journal of Man-Machine Studies. – 1986. – Vol. 24(1). – P. 65-75. –doi: 10.1016/S0020-7373(86)80040-2.
- [Mansouri et al., 2023] Mansouri T., Vadera S. Explainable fault prediction using learning fuzzy cognitive maps // Expert systems. –2023. – Vol. 40(8). – Article e13316. – doi: 10.1111/exsy.13316.
- [Wang et al., 2021] Wang J., Peng Z., Wang X., Li C., Wu J. Deep Fuzzy Cognitive Maps for Interpretable Multivariate Time Series Prediction // IEEE Transactions on Fuzzy Systems. –2021. – Vol. 29(9). –P. 2647–2660. –doi: 10.1109/TFUZZ.2020.3005293.
- [Wang et al., 2022] Wang W., Wang Y., Han X. A dynamic failure mode and effect analysis for train systems failures risk assessment using FCM and prospect theory // Management System Engineering. –2022. –Vol. 1. –Article 8. –doi: 10.1007/s44176-022-00008-x.