

УДК 665.6: 51-7

**ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ  
И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА  
В ХИМИЧЕСКОЙ ТЕХНОЛОГИИ.**

**ЧАСТЬ II**

**USING METHOD OF MACHINE TRAINING AND ARTIFICIAL  
INTELLIGENCE IN CHEMICAL TECHNOLOGY.**

**PART II**

**В.Р. Нигматуллин, Н.А. Руднев**

**Уфимский государственный нефтяной технический университет,  
г. Уфа, Российская Федерация**

**Vil R. Nigmatullin, Nikolay A. Rudnev**

**Ufa State Petroleum Technological University, Ufa, Russian Federation  
e-mail: ngb2008@mail.ru**

**Аннотация.** Статья является второй частью работы, в которой рассмотрено применение методов машинного обучения и искусственного интеллекта для различных задач химической технологии, таких как моделирование, автоматизация и оптимизация процессов, контроль качества и безопасности, поиск новых соединений и катализаторов. Для данных целей были использованы искусственные нейронные сети, метод решающих деревьев, бустинг, регрессия, а также их комбинации.

Во второй части работы рассматриваются следующие вопросы: умное производство, технологии больших данных, поиск и обнаружение неисправностей, предсказание свойств катализаторов, а также кратко представлены наиболее популярные программные средства.

**Abstract.** The article is the second part of the work, which discusses the use of machine learning and artificial intelligence methods for chemical technology various problems, such as modeling, automation and optimization of processes, quality and safety control, search for new compounds and catalysts. For these purposes, artificial neural networks, the method of decision trees, boosting, regression, and their combinations were used.

The second part of the work addresses the following issues: smart manufacturing, big data technologies, search and fault detection, prediction of catalyst properties, as well as the most popular software tools are briefly presented.

**Ключевые слова:** машинное обучение; нейронные сети; глубокие нейронные сети; бустинг; решающие деревья; оптимизация; инструментарии машинного обучения; математическая модель; Big Data

**Key words:** machine learning; neural networks; deep neural networks; boosting; decision trees; optimization; machine learning tools; mathematical model; Big Data

## Введение

Во второй части работы, посвященной анализу основных тенденций при применении методов машинного обучения и искусственного интеллекта в химии и химической технологии, рассмотрены вопросы о внедрении умного производства, технологиях больших данных, поиске неисправностей на основании потока данных на производстве, предсказании параметров катализаторов и дан краткий перечень инструментов, используемых разработчиками информационных систем и программ в этой области.

Внедрение цифровых технологий в химической промышленности связано с рядом особенностей, которых нет например в машиностроении, и

использование опыта исследователей в области применения различных алгоритмов и методов может быть полезно как при принятии решений в производственной деятельности, так и при создании обучающих материалов для студентов вузов.

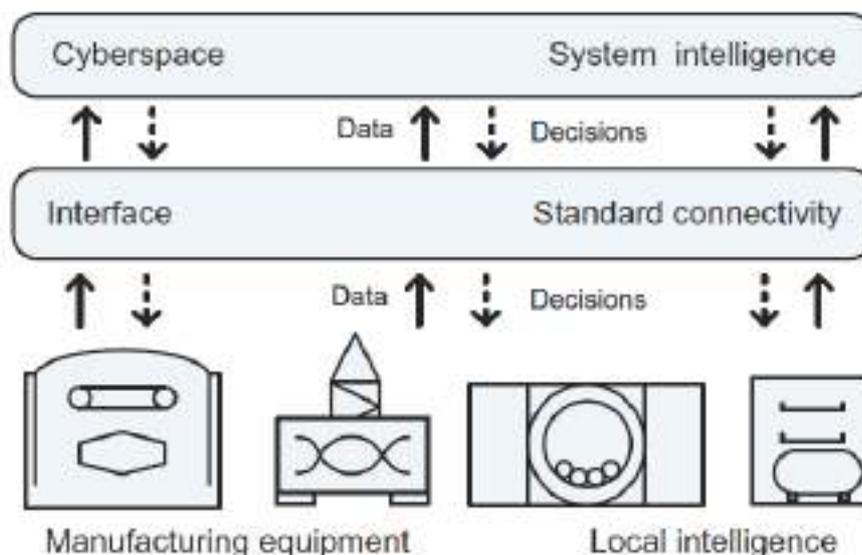
### **Умное производство**

Умное производство использует передовую аналитику данных для дополнения физических законов для повышения эффективности работы системы и принятия решений. При широком распространении датчиков и Интернета вещей (IoT) возрастает потребность в обработке больших производственных данных, характеризующихся высоким объемом, высокой скоростью и высоким разнообразием. Глубокое обучение предоставляет расширенные аналитические инструменты для обработки и анализа больших производственных данных.

Различные страны разработали стратегические планы для преобразования производства, чтобы использовать преимущества новой инфраструктуры, представленной Интернетом вещей и наукой об обработке данных. Например, в 2010 г. Германия представила промышленную систему «Индустрия 4.0», которая теперь развивается всеми странами-членами Европейского союза совместно. Аналогичным образом в 2011 г. Коалиция лидеров Smart Manufacturing (SMLC) в США создала системную основу для внедрения интеллектуального производства.

Поскольку производственные машины все больше оснащены датчиками и средствами связи, существует значительный потенциал для дальнейшего повышения осведомленности о состоянии производственных машин и процессов, сокращения времени простоя в эксплуатации, повышения уровня автоматизации и качества продукции, а также более своевременного реагирования на динамично меняющиеся требования

клиентов [1–6]. Общая концепция умного предприятия показана на рисунке 1.



**Рисунок 1.** Концепция умного предприятия

*Интеллектуальное производство* – это новая производственная парадигма, в которой производственные машины подключены через беспроводные сети, контролируются датчиками и управляются вычислительным интеллектом, и благодаря этому повышаются качество продукции, производительность и устойчивость при одновременном снижении затрат. Внедрение на производстве Интернета вещей, облачных вычислений, киберфизической системы (CPS) способствует развитию современного производства [7–10].

Интеллектуальная обработка данных привела к обширным исследованиям в области прогнозирования данных и принятия решений. Методы интеллектуального анализа данных подразделяются на несколько категорий: классификацию, прогнозирование, кластеризацию и эволюционный анализ. Области применения методов машинного обучения рассмотрены для многих производств, их сильные и слабые стороны также найдены и рассмотрены [11, 12]. Роль управления данными в интеллектуальном производстве представлена на рисунке 2.



**Рисунок 2.** Управление данными в умном производстве

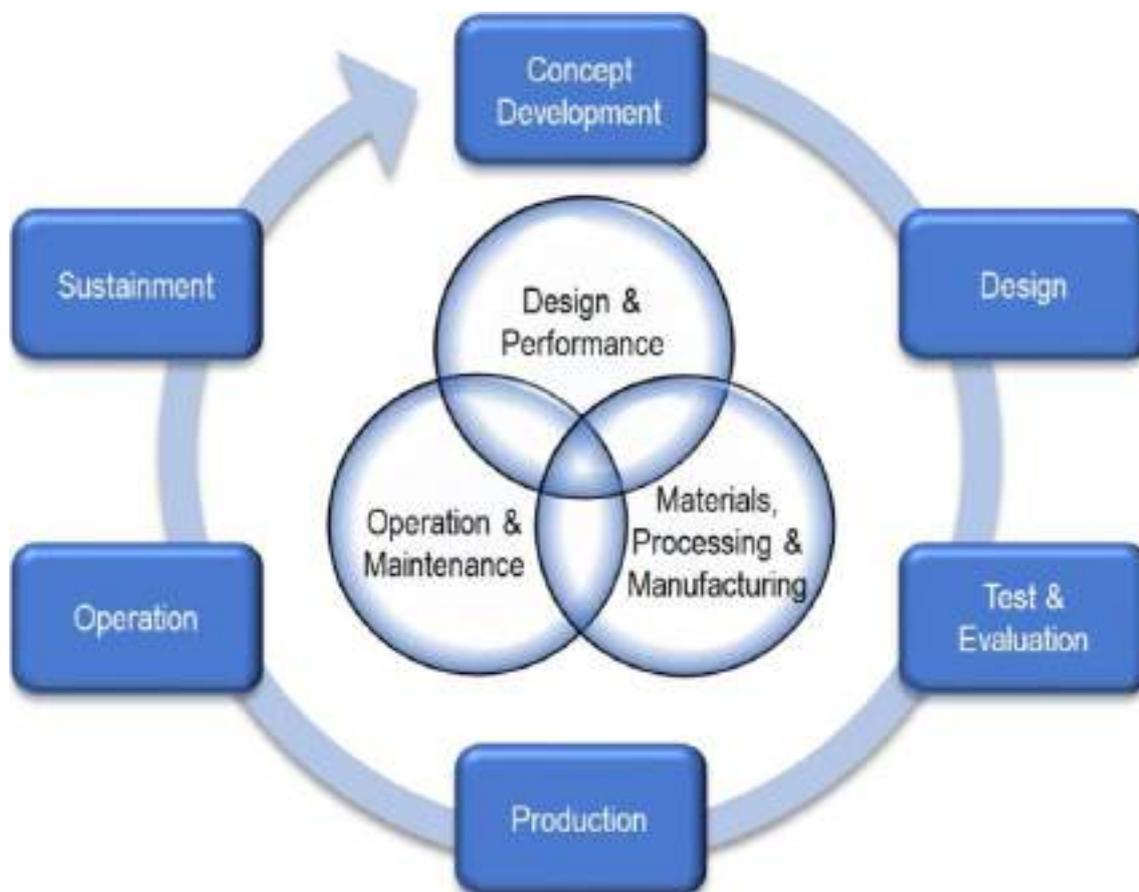
Как прорыв в области искусственного интеллекта, *технология глубокого обучения* демонстрирует выдающуюся производительность в различных приложениях: распознавания речи, восстановления изображения, обработки естественного языка (например, перевода, понимания, тестовых вопросов и ответов). Глубокое обучение обеспечивает расширенный инструмент аналитики для интеллектуального производства в эпоху больших данных. Оно предлагает хороший потенциал для стимулирования производственных приложений, управляемых данными [13, 14]. На рисунке 3 показаны типичные сценарии применения машинного обучения в интеллектуальном производстве.

В 2008 г. Х. Се рассмотрел применение методов машинного обучения для инспекции поверхностной и обработки изображений для выявления дефектов, чтобы повысить качество продукции на производстве [15].

Машинное обучение достигло значительных успехов и дает надежные результаты при оценке поверхностей во многих случаях [16].

Производственные системы также обычно подвержены сбоям, вызванным износом или ненормальными условиями эксплуатации, что приводит к чрезмерной нагрузке, деформации, разрушению, перегреву, коррозии и износу. Отказ может повлечь за собой более высокие эксплуатационные расходы, более низкую производительность,

неоправданные потери деталей и даже непредвиденные простои. Для того, чтобы реализовать разумное производство, крайне важно для условий «умного» фабричного оборудования идентифицировать зарождающиеся дефекты, выявить основную причину отказов, а затем включить информацию в производство и контроль [17].



**Рисунок 3.** Применение методов машинного обучения в умном производстве

Было проведено множество исследований по оценке износа оборудования, все больше и больше методов глубокого обучения исследуются для диагностики и классификации отказов оборудования. *Искусственная нейронная сеть* объединяет функции обучения и диагностики дефектов для одной модели и позже применяется на других схожих деталях, например, такая модель строилась для подшипников, редукторов, ветрогенераторов, роторов [18–25]. Модели, построенные с использованием методов машинного обучения, были использованы при

диагностике неисправностей двигателей самолетов, химических процессов и аппаратов, поршневых компрессоров, подшипников качения [26–28].

В целях повышения производительности при одновременном снижении затрат на техническое обслуживание крайне важно разработать и внедрить такую стратегию, которая позволяет производителям определять состояние систем, находящихся в эксплуатации, для прогнозирования того, когда следует проводить техническое обслуживание. Недавно была исследована *общая рекуррентная нейронная сеть* для прогнозирования распространения дефектов и оценки оставшегося срока полезного использования механических систем или компонентов. Сеть LSTM позволяет изучать характеристики изменяемые во времени и предназначена для прогнозов требующих высокой точности [29, 30].

*Умное производство* – это не степень автоматизации производственного цеха; речь идет об автономии, развитии, моделировании и оптимизации производственного предприятия. Моделирование и оптимизация будут зависеть от доступности данных и инструментов. Уровень «умности» производства будет определяться степенью, в которой физическое предприятие было отражено в киберпространстве.

Глубокое обучение предоставляет большие возможности для умного производства в эпоху Big Data. Путем обработки большого объема данных глубокое обучение дает лицам, принимающим решения, новое видение их операций, а также требуемые показатели в режиме реального времени. Но все же несмотря на многообещающие результаты, о которых сообщалось до сих пор, все еще существуют некоторые ограничения и серьезные проблемы для реализации, такие как плохая техническая оснащенность предприятий и малое количество данных для обучения. По мере развития вычислительных ресурсов обучение может быть реализовано с использованием *облачных технологий* [31, 32].

## Big Data в химической промышленности

Одновременное достижение высокой эффективности, безопасности и прибыльности процесса представляет наибольший интерес в современной обрабатывающей и перерабатывающей промышленности, хотя это и не просто [33].

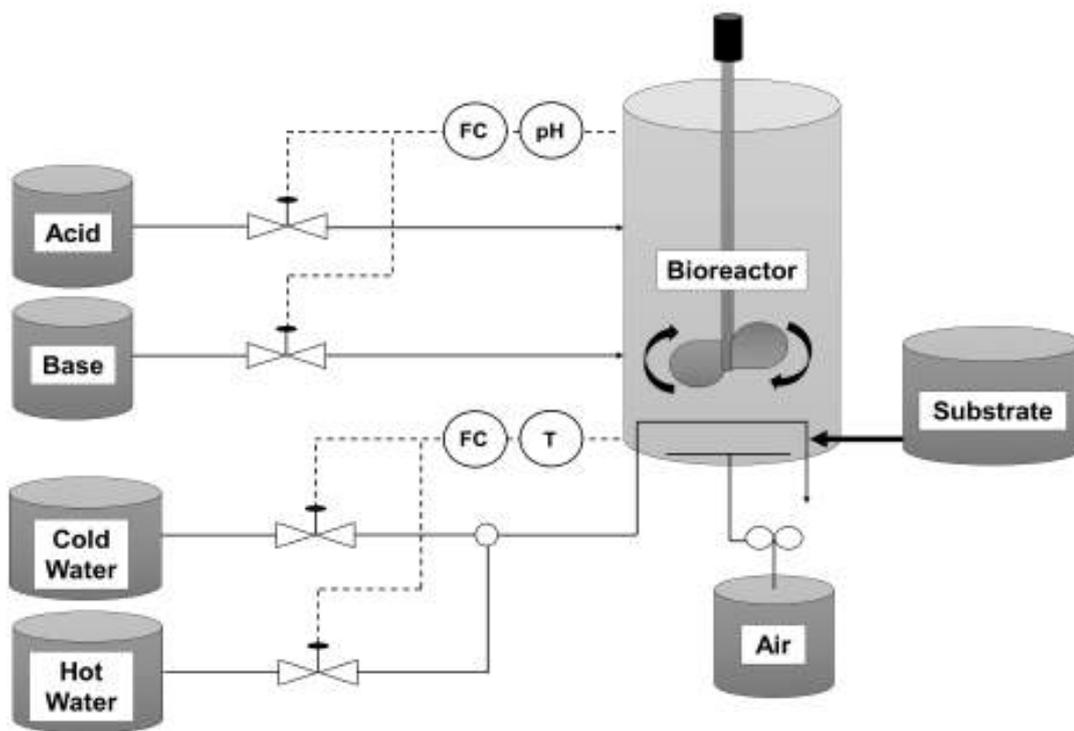
В последнее время умное производство, которое объединяет автоматизированные цифровые технологии с передовыми производственными возможностями [34], вызвало значительный интерес со стороны научных кругов, промышленности и правительства. В результате чего происходит увеличение сбора большого количества данных для всех технологических процессов, часто называемых *большими данными* [35].

Сегодня, используя промышленные большие данные, можно обнаруживать неисправности, диагностировать их по ключевым переменным процесса, прогнозировать будущее состояние переменных процесса и предотвратить любые нежелательные условия [35].

Ошибка при проведении процесса возникает, когда есть недопустимое отклонение хотя бы в одной наблюдаемой переменной или вычисленном параметре системы, и управляющие контроллеры не могут его исправить [36].

Раннее и быстрое обнаружение и диагностика технологических сбоев являются одной из главных задач промышленности для поддержания безопасной эксплуатации и минимизация производственных потерь [37]. Эти проблемы решаются путем мониторинга процесса.

В исследовании 2017 г. М. Онел рассмотрел возможности использования Big Data для получения данных при построении модели для мониторинга и поиска неисправностей технологического процесса производства пенициллина. Схема рассмотренной установки представлена на рисунке 4.



**Рисунок 4.** Схема производства пенициллина

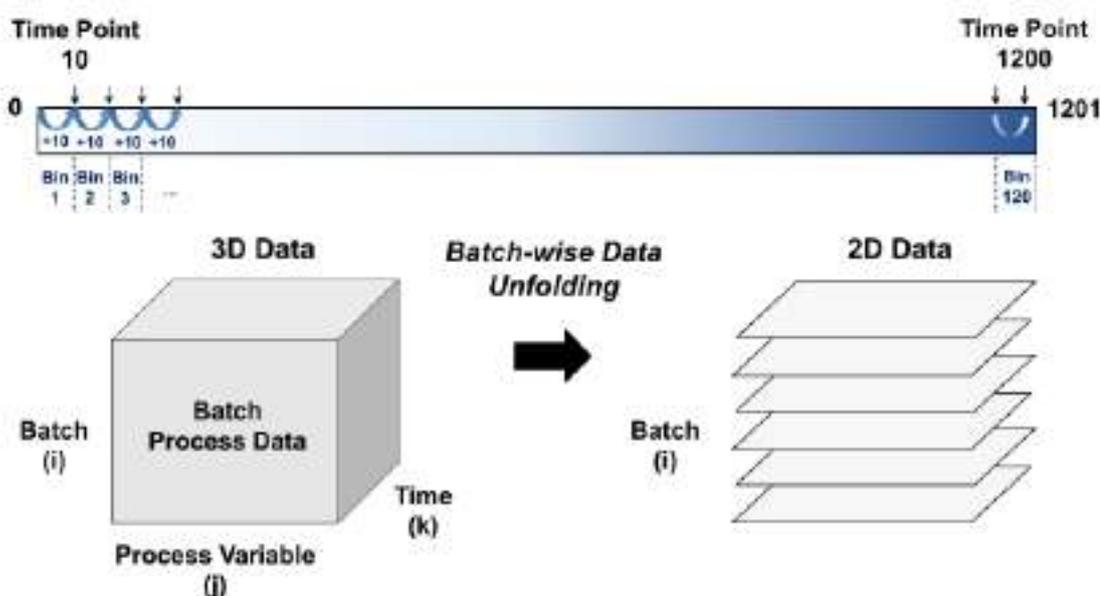
*Мониторинг процессов* с использованием Big Data состоит из двух этапов:

1) *автономный этап* включает в себя разработку моделей для обнаружения и диагностики неисправностей с использованием исторических данных обработки сигналов, в которых используется новый алгоритм выбора признаков, основанный на оптимизации;

2) *оперативный этап* отслеживает текущие партии в режиме реального времени с использованием моделей неисправности и времени [38].

Изначально происходит создание выборочных наборов данных. Для каждой партии в каждый конкретный момент времени собираются все возможные параметры процесса. При сборе данных нам необходимо учитывать время каждого сбоя. Для этого время разбивается на интервалы фиксированного размера, и параметры процесса замеряются в определенные периоды времени. Эти периоды времени называются выборочные наборы.

При автономном этапе цель модели состоит в том, чтобы принять решение об обнаружении и диагностике неисправностей в онлайн-режиме. Чтобы обучить модель, нужно преобразовать данные в двухразмерную матрицу, как показано на рисунке 5.



**Рисунок 5.** Получение данных процесса и формирование образцов обучения

После создания и подключения к процессу модели следует фаза обнаружения и диагностики неисправностей. На этом этапе создается инструмент поддержки принятия решений для оперативного обнаружения и диагностики неисправностей. Это выбранные бинарные классификаторы, оптимальное решение функции, которые будут оценивать поступающие, предварительно обработанные данные о процессе и давать бинарный ответ о возникновении ошибки.

*Модели классификаторов ошибок* включают оптимальный набор характеристик, которые являются наиболее информативными для процесса, чтобы возможно было диагностировать обнаруженную неисправность в интересующий период времени [38].

Чтобы работать с анализом больших данных, аппаратная платформа всего предприятия может быть дополнена кластером

высокопроизводительных рабочих станций. Однако эта стратегия дорога и неустойчива. В качестве альтернативы – если можно использовать как распределенные, так и параллельные стратегии для развертывания хранения и анализа больших данных на обычных персональных компьютерах, то стоимость может быть значительно снижена, а масштабируемость может быть значительно улучшена [39].

С другой стороны, в связи с развитием промышленности 4.0 также требуется, чтобы производственная система была интеллектуальной, интегрированной и эффективной для содействия принятию решений [40]. С этой целью парадигма мониторинга в виде интеллектуального модуля должна обеспечивать многоуровневую интегрированную визуализацию, которая приведет к быстрому и надежному устранению неисправностей в крупных системах.

Apache Hadoop – это проект с открытым исходным кодом для хранения и анализа больших данных [41]. Apache Hadoop содержит два основных компонента: распределенную файловую систему Hadoop (HDFS) и MapReduce. В частности, HDFS предоставляет распределенную, масштабируемую и сбалансированную платформу хранения, которая высоко оптимизирована для высокопроизводительных операций с данными. Для вычислений MapReduce представляет собой пакетно-ориентированную, распараллеленную и отказоустойчивую среду анализа с кластером вычислительных узлов. На основе этой схемы многие алгоритмы интеллектуального анализа данных и машинного обучения, такие как логистическая регрессия, случайный лес и метод опорных векторов, могут быть адаптированы и распараллелены для анализа больших данных [42].

Как правило, Hadoop MapReduce предлагает среду параллельного программирования, которая позволяет разработчикам распределять огромные объемы данных определенным образом по кластеру недорогих вычислительных узлов [41].

В проведенном в 2018 г. исследовании использовалась модель для мониторинга крупномасштабных промышленных процессов с несколькими условиями эксплуатации. На основе платформы Hadoop MapReduce были разработаны как алгоритмы параллельной стандартизации данных, так и алгоритмы обучения смешанной модели для решения проблемы больших данных методом «разделяй и властвуй». Впоследствии была разработана иерархическая стратегия мониторинга с целью предоставления диаграммы мониторинга для анализа на уровне всего предприятия или отдельных его блоков. В ходе эксплуатации модель показала высокую эффективность и стабильную работу [43].

Для полной оптимизации предприятия и построения вспомогательного механизма для принятия решений требуется собрать и обработать большое количество данных. В настоящее время это возможно сделать только с использованием технологий Big Data и распределенных файловых систем по типу Hadoop.

### **Обнаружение неисправностей и безопасность**

Компании способны собирать жизненно важные данные о процессах на разных этапах производства. При использовании доступных данных имеется возможность использовать аналитику данных, чтобы позволить компаниям строить модели для обнаружения неисправностей и контроля безопасности.

Недавние исследования в области машинного обучения для моделирования производственных областей с особым упором на профилактическое обслуживание и обнаружение неисправностей доказали, что прогнозы редких событий и создание систем прогнозирования неисправностей с использованием альтернативных методов возможно и дает достаточно точные результаты [44–46].

Ван и его коллеги продемонстрировали использование баесовских сетей (BN) для моделирования обнаружения неисправностей в процессе

производства полупроводников, где присутствуют как дискретные, так и непрерывные параметры. Используя знания экспертов, были определены шесть переменных для мониторинга состояния исследуемого процесса. Хотя это и продемонстрировало потенциал BN, модель была точной только при использовании адекватных обучающих данных, содержащих потенциальные отклонения.

В исследовании 2018 г. корректность работы модели, построенной с использованием BN была проверена на данных, взятых в компании Bosch. Было выявлено, что с увеличением обучающей выборки модель точно прогнозирует даже непрерывные параметры рассмотренных процессов [47].

С широким применением распределенных систем управления (DCS) процессы становятся все более автоматизированными. Но все же было множество аварий, связанных с химическими процессами и аппаратами и приведших к жертвам, а также к ущербу окружающей среде.

До сих пор один критический уровень защиты на предприятиях все еще отсутствует. Эта задача в настоящее время выполняется операторами, которые должны своевременно знать о неисправностях и принимать корректирующие решения. Для операторов, которым не хватает квалификации либо на которых действуют мешающие работе факторы, а также при нехватке данных определение неисправностей и угроз практически невозможно. Поэтому возникла промышленная необходимость разработки интеллектуальной системы обнаружения и диагностики неисправностей (FDD), чтобы помочь операторам в обработке ненормальных ситуаций в ходе эксплуатации оборудования.

Методы FDD могут быть разделены на статистические методы и методы глубокого обучения. Статистические методы включают принцип компонентного анализа (PCA), независимого компонентный анализа (ICA), метод наименьших квадратов (PLS), анализ Фишера (FDA) [48, 49].

Методы обучения включают в себя методы опорных векторов (SVM), k-ближайших соседей (KNN), искусственные нейронные сети (ANN). Эти методы обучения были успешно использованы для диагностики неисправностей. Кроме того, нейронные сети уже применяются при разработке моделей FDD, такие как HANN, Duty-Oriented HANN (DOHANN) и локальный многослойный персептрон (SLMLP) [50, 51].

Неисправности в химических процессах – это, по сути, заявления о том, что технологические переменные отклоняются от их нормального состояния.

Данные из различных состояний отклонения могут быть использованы для диагностики типов ошибок.

Большинство существующих методов FDD рассматривают только признаки неисправности, но не определяют их.

В 2018 г. глубокие нейронные сети были рассмотрены для построения алгоритма определения неисправностей и отклонений. На рисунке 6 представлена основа метода диагностики неисправностей на основе использования глубоких нейронных сетей (DCNN) [52].

В результате эксперимента было доказано, что метод диагностики неисправностей, основанный на DCNN, имеет отличную производительность. Точность метода достигает 88,2 %, что выше, чем результат диагностики с использованием других методов, опубликованных в литературе. Если отбросить сложно диагностируемые ошибки, то можно повысить точность до 91 % [52].

Метод DCNN перспективен для промышленного применения из-за его высокой точности диагностики. Тем не менее, модель по-прежнему опирается на исторические данные о неисправностях. Также для сверхсложного химического процесса, который обычно имеет более тысячи параметров, количество входных данных будет очень велико, это создает определенные проблемы при разработке архитектуры DCNN.

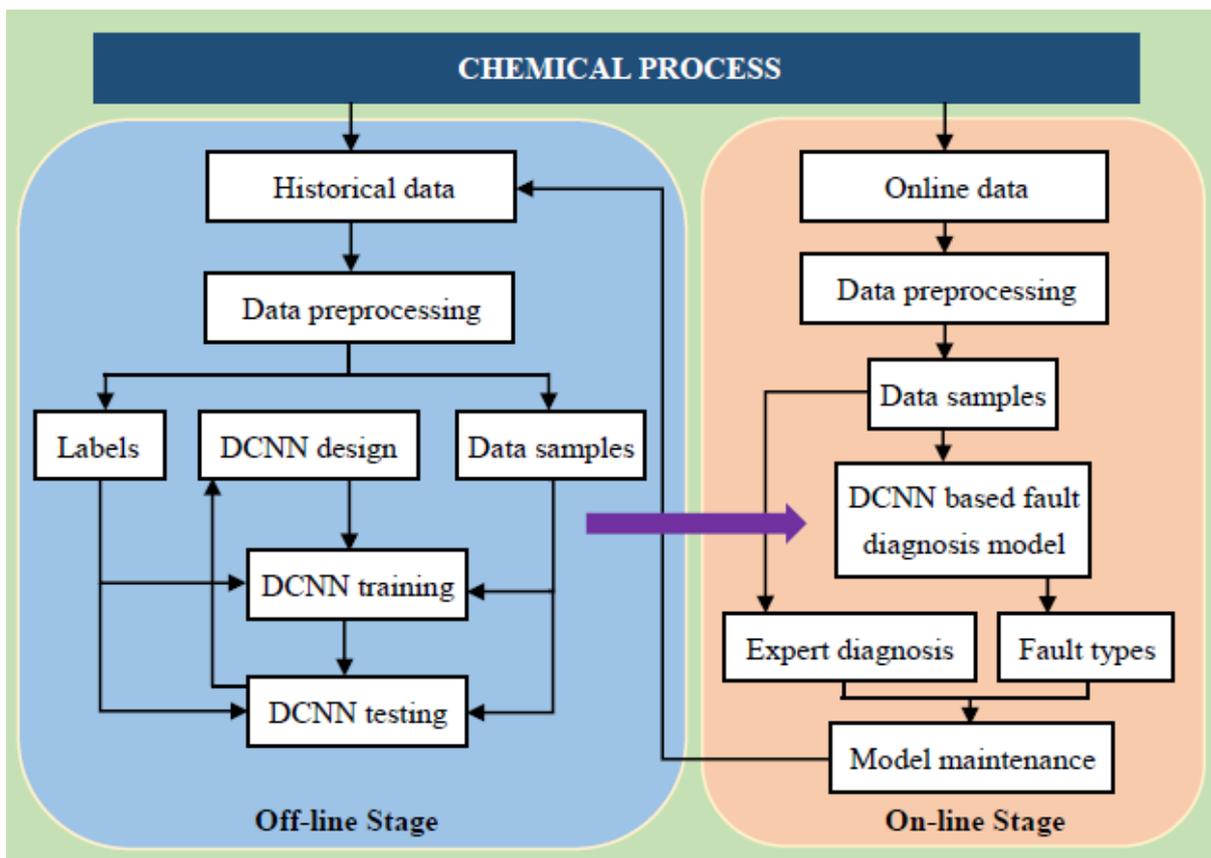


Рисунок 6. Метод диагностики неисправностей с применением нейронных сетей

### Катализаторы

По сравнению с другими областями науки применение методов машинного обучения в области катализа не было хорошо изучено в течение последних десятилетий по следующим двум причинам: отсутствие достаточно большой базы данных (как экспериментальные, так и теоретические исследования каталитических реакций требуют большого количества времени и ресурсов), слишком много входных переменных (чтобы предсказать свойства реакции, обычно есть большое количество независимых переменных, которые коррелируют с предсказанными свойствами, например, условия эксперимента, координаты реакции, электронные структуры и орбитальные свойства). Эти проблемы значительно затрудняют разработку и применение машинного обучения для каталитических процессов.

Для катализа с использованием наноразмерных катализаторов все становится еще сложнее: экспериментальные характеристики их катализаторов сложны, и есть много наноразмерных эффектов, что трудно наблюдать. Теоретически, хотя есть некоторые современные вычислительные методы (например, квантово-механические вычисления), которые могут точно рассчитать электронную структуру и энергетические свойства каталитической системы (например, наночастиц и наностержней), высокие вычислительные затраты – все еще огромная проблема [53].

Но в последнее десятилетие быстрое развитие алгоритмов компьютерного и машинного обучения показывает, что некоторые методы машинного обучения могут выполнить быстрое обучение даже с использованием достаточно большой базы данных для обучающей выборки.

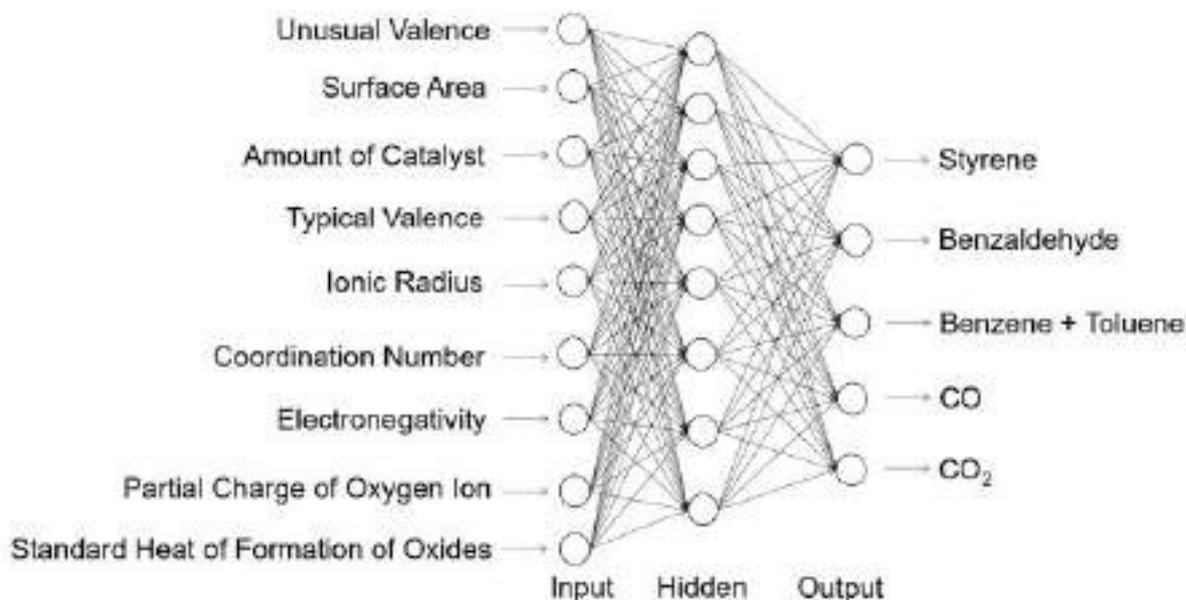
Тем не менее, существует очень мало исследований, которые обобщают приложения машинного обучения для моделирования катализа. Поэтому подробный анализ применения ANN для экспериментального и теоретического катализа очень важен [53].

Самым простым применением ANN является численный прогноз. Одна из самых ранних работ по применению машинного обучения в приложении к катализу была проведена в 1994 г. В этой работе предсказано распределение продукта этилбензола в процессе окислительного гидрирования, с компонентами продукта в качестве выходов сети.

Для данных, подаваемых в качестве входных в ANN, были использованы девять различных независимых параметров, которые имели отношение к производительности и селективности каталитической реакции, такие как площадь поверхности катализатора, количество катализатора, валентность, координационное число, электроотрицательность и стандартная теплота образования оксидов. Результаты показали, что с применением хорошей экспериментальной

базы данных модель ANN может выполнять точные прогнозы для состава продукта.

Схематическая структура ИНС, которая использована для прогноза, реконструирована на рисунке 7 [54].



**Рисунок 7.** Прогнозирование каталитической активности этилбензола

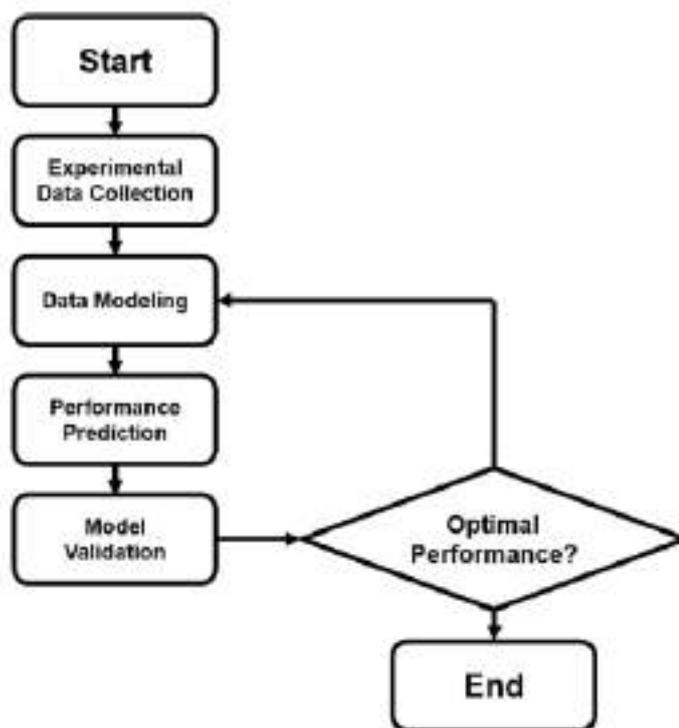
Для более сложных реакций, таких как эпексидирование 1-гексена катализируемого комплексами Мо (VI) на полимерной подложке, в ходе проведенного в 2012 г. исследования было доказано, что применение ANN имеет мощный потенциал для прогнозирования каталитической активности, данные, полученные в качестве прогноза, согласовывались с экспериментальными выводами [55].

Помимо катализа, выяснилось, что подобная концепция может даже помочь открытию материалов по неудачным экспериментальным данным [56].

В дополнение к прогнозированию активности катализатора ученые начали задумываться еще над одним вопросом – как можно проектировать новые катализаторы, используя предсказательную силу ANN.

Когда известно, что ИНС могут точно предсказать свойства различных каталитических систем, возможно разработать и сгенерировать модели на основе ANN для получения новых ожидаемых.

Общая алгоритмическая схема оптимизации катализатора, обобщенная Мальдонадо и Ротенбергом, представлена на рисунке 8.



**Рисунок 8.** Оптимизационное моделирование

Для более точных результатов при оптимизации процесса аммоксидирования пропана исследователи объединили ИНС с генетическими методами (GA), что ускорило выбор катализатора. Точно также Умегаки объединили GA и ANN для оптимизации оксидного катализатора Cu-Zn-Al-Sc для синтеза метанола. Родемерк обобщил метод ANN с помощью GA и предложил общую основу для скрининга новых твердых каталитических материалов, эта методика хорошо согласуется с экспериментальными данными.

На основе предыдущих разработок методов ANN и GA Баумес развил идею «ANN filter» для высокопроизводительного скрининга (HTS) гетерогенного катализа [57]. На примере реакции конверсии водяного пара

(WGS) было показано, что ранее применяемые методы не смогли точно оценить активность реакции WGS, однако хорошо обученный классификатор ANN-фильтр может помочь определить «хорошие» и «плохие» катализаторы.

ANN хорошо подходит для прогнозирования каталитической активности. В ходе многочисленных опытов доказано, что ИНС, обученная на достаточно большой базе данных, может выполнять прогноз и оптимизацию процесса катализа с высокой точностью. Для повышения эффективности при проектировании новых катализаторов самой успешной стратегией является совмещение ANN и GA.

### **Инструментарий для разработки систем искусственного интеллекта**

В работе проведен анализ публикаций и инструментов, применяемых области машинного обучения и искусственного интеллекта. Самые крупные ИТ компании в отрасли (Google, Microsoft, Apple, Samsung, IBM, Nvidia, Intel и прочие) активно включились в соревнование в области создания, применения и разработки решений и инструментария в этой области.

Растет число программных инструментов, доступных разработчикам для проектирования сложных систем. Различные структуры ИИ и API-интерфейсы упрощают внедрение машинного обучения в разрабатываемые проекты.

Надо отметить, что разработчикам доступно большое число инструментов, свободно распространяемых и открытых исходным кодом.

Accord.NET Framework – это среда машинного обучения, объединенная с библиотеками обработки аудио и изображений [58].

H2O – это программный инструмент с открытым исходным кодом [59].

Apache PredictionIO – это сервер машинного обучения с открытым исходным кодом [60].