

Министерство науки и высшего образования Российской Федерации
Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа прикладной математики и вычислительной физики

Работа допущена к защите
Руководитель ОП
_____ А.М. Левченя
« ____ » _____ 2023г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(работа бакалавра)

**ОПЫТ ПРИМЕНЕНИЯ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСЧЕТА
ДВУМЕРНЫХ ТЕЧЕНИЙ ВЯЗКОЙ ЖИДКОСТИ**

по направлению подготовки 03.03.01 – Прикладные математика и физика
Направленность (профиль) 03.03.01_01 – Математические модели и
вычислительные технологии в гидроаэродинамике и теплофизике

Выполнил
студент гр. 5030301/90101

М.С. Трошин

Руководитель
доцент, к.ф.-м.н.

А.А. Смирновский

Санкт-Петербург
2023

РЕФЕРАТ

На 49 с., 44 рисунка, 5 таблиц.

КЛЮЧЕВЫЕ СЛОВА: МАШИННОЕ ОБУЧЕНИЕ, ФИЗИКО-ИНФОРМИРОВАННАЯ НЕЙРОННАЯ СЕТЬ, ЛАМИНАРНОЕ ТЕЧЕНИЕ, КАНАЛ С ВНЕЗАПНЫМ СУЖЕНИЕМ, PINN, TENSORFLOW, PYTORCH.

Тема выпускной квалификационной работы: «Опыт применения нейронных сетей для расчета двумерных течений вязкой жидкости».

Данная работа посвящена исследованию возможностей применения нейронных сетей в области гидродинамики с целью получения достоверных гидродинамических характеристик.

Задачи, которые решались в ходе исследования:

1. Обзор и анализ существующей литературы, посвященной применению нейронных сетей в гидродинамике.
2. Исследование и выбор оптимального подхода к прогнозированию.
3. Реализация программы для моделирования ламинарного течения в канале с резким сужением.
4. Проведение серии вычислительных экспериментов на основе разработанной программы для различных чисел Рейнольдса.
5. Разработка нейронной сети с обратным прогнозом.

В данном исследовании было выбрано применение физико-информированной нейронной сети, ввиду их эффективности в сценариях прогнозирования, основанных на физических законах и ограничениях. Отличительная особенность данных сетей заключается в возможности снижения зависимости от исходного объема данных или даже его полного исключения в ходе обучения нейронной сети. Весь процесс проектирования и разработки проводился с применением открытых библиотек и языка программирования Python.

В результате проведенной работы получены ценные навыки и знания в области применения нейросетевых методов в гидродинамике. Были изучены и применены актуальные инструменты для разработки и проектирования, а также получены результаты, отражающие современные достижения в области использования технологии нейронных сетей в вычислительной гидродинамике.

ABSTRACT

49 pages, 44 figures, 5 tables.

KEYWORDS: MACHINE LEARNING, PHYSICS-INFORMED NEURAL NETWORK, LAMINAR FLOW, SUDDEN CONSTRICTION CHANNEL, PINN, TENSORFLOW, PYTORCH.

The subject of the graduate qualification work is «Experience in using neural networks for calculating two-dimensional flows of viscous fluid».

This work is dedicated to exploring the possibilities of applying neural networks in the field of hydrodynamics to obtain reliable hydrodynamic characteristics.

The tasks that were solved in the course of the research are:

1. Review and analysis of existing literature dedicated to the application of neural networks in hydrodynamics.
2. Exploration and selection of the optimal approach to prediction.
3. Implementation of a program for modeling laminar flow in a channel with a sharp narrowing.
4. Conducting a series of computational experiments based on the developed program for different Reynolds numbers.
5. Implementation of an Inverse neural network.

In this study, the use of physics-informed neural networks was chosen due to their effectiveness in prediction scenarios based on physical laws and constraints. A distinctive feature of these networks is the possibility of reducing dependence on the initial volume of data or even its complete exclusion during the training of the neural network. The entire design and development process was carried out using open libraries and the Python programming language.

As a result of the work carried out, valuable skills and knowledge were gained in the field of application of neural network methods in hydrodynamics. Actual tools for development and design were studied and applied, and results were obtained that reflect modern achievements in the field of using neural network technology in computational hydrodynamics.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1. ОБЗОР ЛИТЕРАТУРЫ	7
1.1 Глубокие нейронные сети (Deep Neural Networks, DNN)	7
1.2 Генетические алгоритмы (GA) в сочетании с глубокими нейронными сетями (DNN)	8
1.3 Генеративно-состязательные сети (GAN)	10
1.4 Сверточные нейронные сети (CNN)	11
1.5 Dynamic (Recurrent) RNN	12
1.6 Физико-информированные нейронные сети (PINN)	15
2. ПОСТАНОВКА ЗАДАЧИ, МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ПРЕДВАРИТЕЛЬНЫЕ РАСЧЕТЫ ДЛЯ ЗАДАЧИ ТЕЧЕНИЯ В КАНАЛЕ С ВНЕЗАПНЫМ СУЖЕНИЕМ	20
2.1 Математическая модель	20
2.2 Выбор расчетной области и сетки для проведения эталонных расчетов в пакете Fluent	21
3. УСТРОЙСТВО PINN И ИСПОЛЬЗУЕМЫЕ ИНСТРУМЕНТЫ	25
3.1 Архитектура и принцип работы физико-информированной нейронной сети	25
3.2 Тестирование разработанной программы	27
4. РЕЗУЛЬТАТЫ РАБОТЫ PINN И ИХ ОБСУЖДЕНИЕ	29
4.1 Описание процесса обучения сети	29
4.2 Подбор гиперпараметров	30
4.2.1 Выбор функции активации	30
4.2.2 Выбор инициализатора	31
4.2.3 Выбор оптимального числа нейронов и скрытых слоев	32
4.2.4 Выбор оптимизатора	34
4.2.5 Выбор коэффициента β при граничных условиях	34
4.2.6 Выбор сетки обучения	35
4.3 Сравнение результатов работы PINN и Fluent	36
5. ОБРАТНЫЙ ПРОГНОЗ PINN	41
5.1 Постановка задачи	41
5.2 Структура нейронной сети	44
ЗАКЛЮЧЕНИЕ	47
СПИСОК ЛИТЕРАТУРЫ	48

ВВЕДЕНИЕ

Основной задачей вычислительной гидродинамики является применение уравнений движения жидкости к различным задачам и условиям, для которых аналитические решения часто отсутствуют. Проблема поиска решения уравнений динамики жидкости привела к созданию множества различных численных методов гидродинамических расчетов, но и на данный момент возникают сложности и ограничения, связанные со сложностью решаемых уравнений, необходимостью использования большого числа вычислительных ресурсов. В связи с этим, нейронные сети и машинное обучение стали рассматриваться как один из перспективных подходов решения возникающих проблем.

Первые эксперименты, связанные с применением нейросетей в решении задач гидродинамики, были проведены в 90-е годы прошлого века. В это время был разработан и активно начал применяться механизм обратного распространения ошибки. Этот период является важным этапом в развитии нейросетей.

Первые применения были направлены на анализ потоков, в частности, на классификацию траекторий для велосиметрии слежения за частицами (PTV) и велосиметрии изображения частиц (PIV) (Тео и др., 1991; Grant & Pan, 1995), а также для определения фазовых конфигураций в многофазных потоках (Bishop & James, 1993).

Интересный подход к проблематике был предложен в работе Бальди и Хорника (1989 год), где установлена связь между методом главных компонент (POD) и линейными нейросетями (Baldi & Hornik, 1989). Она была использована для реконструкции турбулентных полей и потоков в пристенной области канала, при этом использовалась информация только о стенках (Milano & Koumoutsakos, 2002). В рамках этого исследования было введено несколько слоев нейронов для улучшения результатов сжатия, что, вероятно, стало первым примером применения глубокого обучения, так как мы его знаем сегодня, в области механики жидкости. Цитирование произведено в соответствии с работой [1].

В последние годы мы становимся свидетелями возрождения интереса к использованию методов машинного обучения в механике жидкости. Подобный интерес в большой степени вызван выдающимися результатами, демонстрируемыми архитектурами глубокого обучения, способными извлекать информационно значимые признаки из представленных данных. Данный подход позволил достичь значительных успехов в различных областях, где данные и ограничения моделей играют важнейшую роль, включая социальные науки, а также в бизнесе, где точность прогнозирования имеет первостепенное финансовое значение. Перспектива совмещения фундаментальных принципов и подходов, основанных на анализе данных, представляется уникальной и, безусловно, обретет своё место в гидродинамике.

Этот синтез традиционных подходов и передовых технологий расширяет горизонты исследований в области гидродинамических процессов. Это особенно важно в рамках прикладных задач, в которых учитывается множество физических аспектов и где полноценные CFD расчеты могут быть крайне ресурсоемкими. К примеру, одной из таких областей является моделирование процессов выращивания кристаллов, где использование физико-информированных нейронных сетей может значительно облегчить работу исследователей.

Таким образом, целью данной работы является проведение анализа научной литературы с целью изучения применения различных типов нейронных сетей в области гидродинамики и моделирования роста кристаллов. Следующим этапом становится отбор и использование определенного типа нейронных сетей для решения базовых гидродинамических задач. Что необходимо для того, чтобы освоить методику их применения, а также уяснить возможные преимущества и недостатки данного подхода.

1. ОБЗОР ЛИТЕРАТУРЫ

В контексте данной работы сфокусируемся на подходах, использующих нейронные сети, и проанализируем ключевые современные примеры их применения в области вычислительной гидродинамики и синтезе кристаллов, в частности.

1.1 Глубокие нейронные сети (Deep Neural Networks, DNN)

Глубокие нейронные сети (DNN) являются основным и наиболее распространенным типом нейронных сетей. Их уникальность заключается в использовании множественных слоев нейронов между входом и выходом сети, что позволяет обнаруживать и предсказывать чрезвычайно сложные зависимости при условии наличия значительного объема данных для обучения. Это делает глубокие нейронные сети многофункциональными и применимыми для широкого диапазона задач, включая классификацию, регрессию и прогнозирование.

Примером успешного использования глубоких нейронных сетей является исследование, представленное в [8]. Используя данные из 450 экспериментов с 100 параметрами в каждом в контексте процесса Чохральского, исследователи успешно разработали и обучили нейронную сеть, которая способна предсказывать целевые параметры на основе входных данных. В работе [7] удалось добиться реализации прогнозирования в реальном времени, используя данные 20 000 временных интервалов для каждого эксперимента, результаты можно наблюдать на рисунке 1.

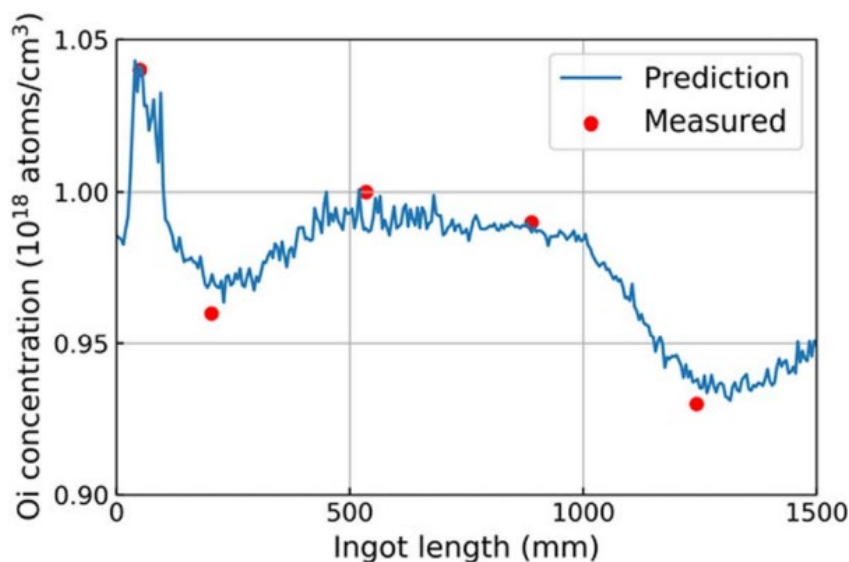


Рисунок 1. [7] Измеренные и прогнозируемые концентрации O_i в одном слитке как функция длины слитка. В системе прогнозирования в реальном времени линия прогноза расширяется вправо по мере роста слитка, и мы можем получить непрерывный профиль концентрации O_i до текущей границы роста в выращенном слитке.

Кроме того, примечательной являются работы [6, 9], где при наличии 1000 CFD расчетов авторам удалось обучить нейронную сеть для предсказания уровня пресыщения и скоростей в зависимости от заданных параметров, результат и постановку задачи можно наблюдать на рисунке 2.

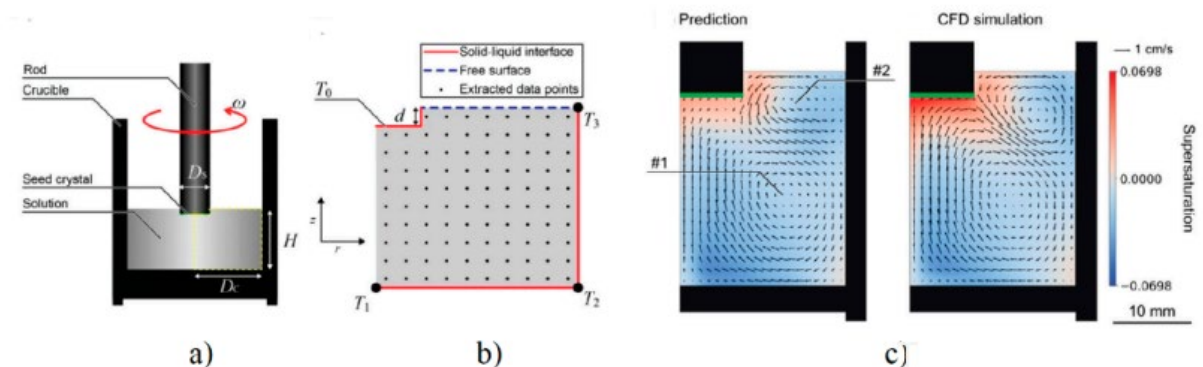


Рисунок 2. [6] Высокоскоростное прогнозирование CFD-симуляций пересыщения и полей скоростей при росте раствора SiC с верхней затравкой (TSSG) с помощью ИНС с 4 скрытыми слоями: (а) конфигурация процесса TSSG, (б) вычислительная область в SiC для CFD, (с) пересыщение и распределение скоростей, предсказанные ИНС (слева) и CFD (справа).

1.2 Генетические алгоритмы (GA) в сочетании с глубокими нейронными сетями (DNN)

GA, или генетические алгоритмы, вдохновлены биологическим процессом естественного отбора и используются для оптимизации. Они работают путем создания популяции кандидатов на решения, которые эволюционируют на протяжении множества поколений с использованием операций, таких как мутация, кроссовер (или скрещивание) и отбор. Эти алгоритмы часто применяются в задачах, где пространство решений велико и трудно обследуемо с помощью классических методов оптимизации.

Обученные глубокие нейронные сети (DNN) часто применяются в качестве средства генерации данных для генетических алгоритмов, тем самым внося значительный вклад в процесс оптимизации.

Примером применения такого совмещенного подхода является исследование [10], где обученная нейросеть использовалась в процессе работы генетического алгоритма для оптимизации формы границы кристалл/сплав и снижения концентрации кислорода на поверхности, результат работы можно увидеть на рисунке 3.

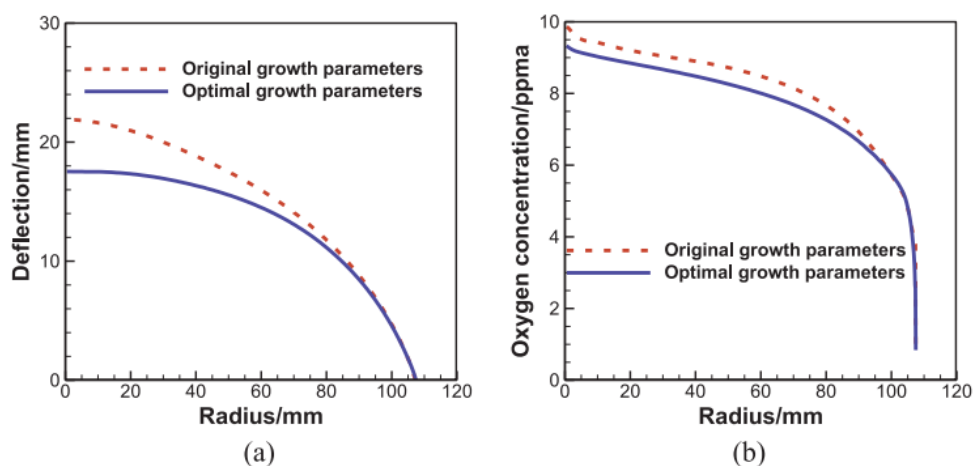


Рисунок 3. [10] Сравнение исходного и оптимального вариантов: (а) форма границы раздела расплав/кристалл и (b) концентрация кислорода вдоль границы раздела.

В работе [11] были успешно применены обученная нейросеть и генетический алгоритм для оптимизации параметров в процессе направленного затвердевания с затравкой квазиоднокристаллического кремния. Это привело к сглаживанию границы раздела затвердевания, снижению теплового напряжения в кристалле и сокращению времени роста, сравнение получаемых полей, можно наблюдать на рисунках 4 и 5.

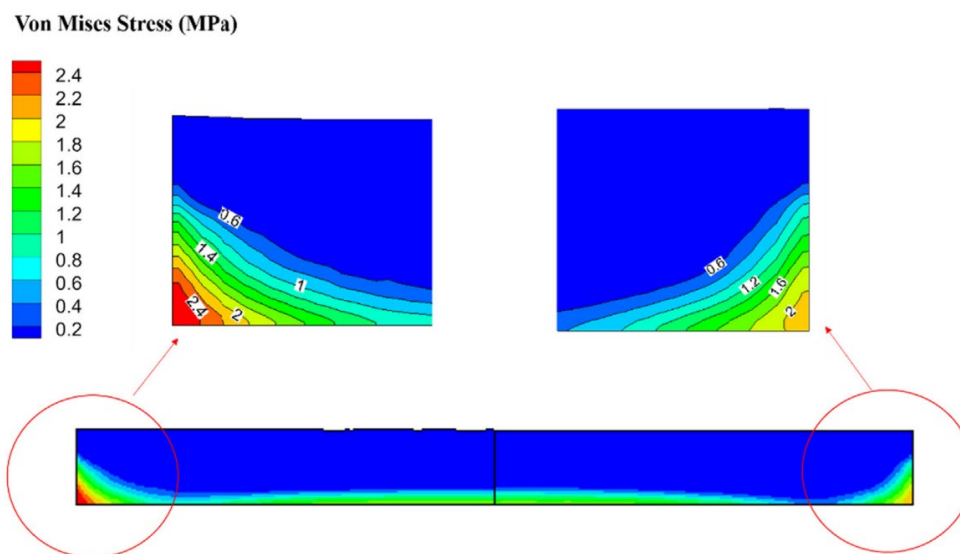


Рисунок 4. [11] Сравнение поля тепловых напряжений в выращенных кристаллах в конце третьей стадии роста между исходными параметрами(слева) и оптимальными(справа).

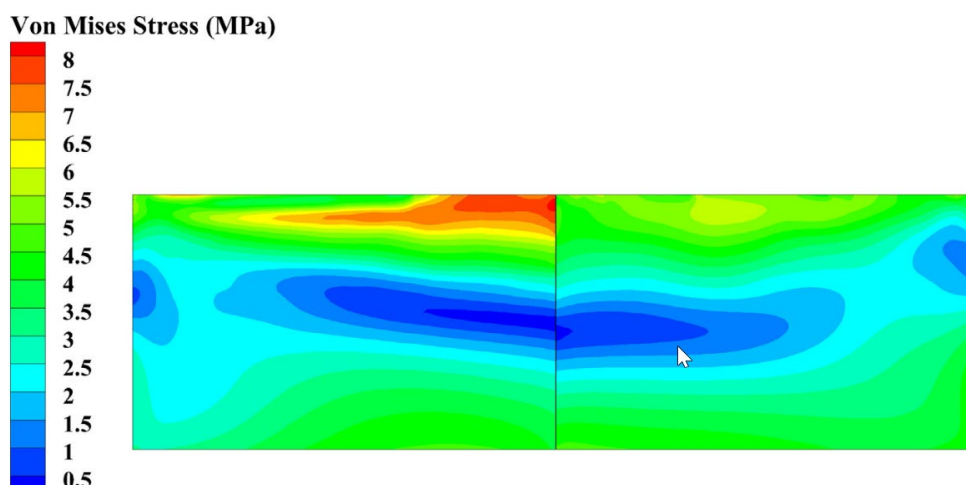


Рисунок 5. [11] Сравнение поля тепловых напряжений в выращенных кристаллах в конце третьей стадии роста между исходными параметрами(слева) и оптимальными(справа).

1.3 Генеративно-состязательные сети (GAN)

Генеративно-состязательные сети (GAN) представляют собой класс машинного обучения, в основе которого лежит идея соревнования двух нейронных сетей: генеративной и дискриминативной. Генеративная сеть стремится создать данные, максимально приближенные к реальным, в то время как дискриминативная сеть старается отличить настоящие данные от сгенерированных. В результате их взаимодействия генеративная сеть обучается создавать данные, которые становятся все более реалистичными.

В контексте гидродинамики GAN могут быть использованы для расширения имеющегося набора данных, улучшения разрешений и моделирования реалистичных гидродинамических явлений. Это открывает новые перспективы для исследований, которые могут быть осуществлены с помощью ограниченного объема экспериментальных данных.

В работе [12] представлено исследование, посвященное анализу различных типов GAN для предсказания полей скорости при обтекании аэродинамического профиля турбулентным потоком. Использование GAN в этом контексте предлагает перспективные подходы к моделированию и прогнозированию сложных турбулентных потоков. Сравнение значений, получаемых с помощью GAN и принятых за эталонные (GT) можно наблюдать на рисунке 6.

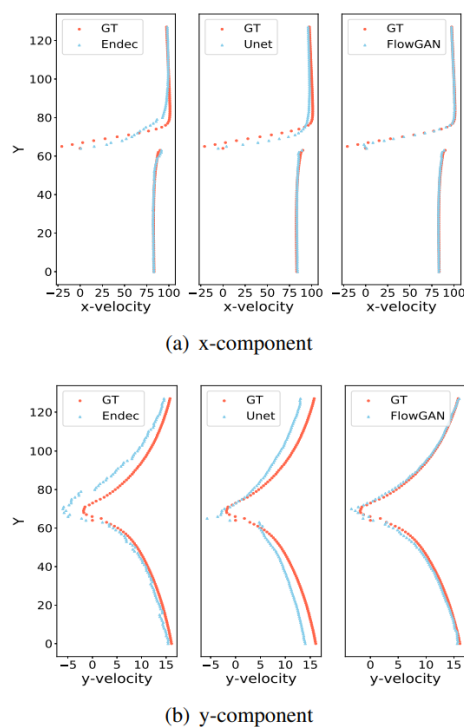


Рисунок 6. [12] Прогнозирование профиля скорости x- и y-компоненты в хвостовой части аэродинамического профиля bw3 с использованием различных моделей в сравнении с наземной истиной (GT).

1.4 Сверточные нейронные сети (CNN)

Сверточные нейронные сети (CNN) представляют собой класс глубоких нейронных сетей, разработанный специально для обработки пространственных данных, таких как изображения. Особенностью этого типа сетей является использование сверточных слоев, которые автоматически и адаптивно учат пространственные иерархии признаков.

Используя свойство сохранения пространственной информации в процессе обучения, CNN эффективно работают с пространственно-структурированными данными, что делает их идеально подходящими для анализа изображений. Это находит свое применение в области вычислительной гидродинамики. Например, CNN могут быть использованы для анализа визуальных представлений параметров течения, полученных в результате численного моделирования, таких как поля скоростей, давления или плотности. Кроме того, на основе предыдущих состояний течения, CNN могут предсказывать будущее, что актуально в задачах прогнозирования и оптимизации. Также с помощью CNN возможно восстанавливать полные данные о течении на основе неполных или неточных измерений.

В работе [13] используется сверточная нейронная сеть для анализа давления в потоке на основе изображения поля скорости. При этом применение CNN позволило учесть пространственные зависимости и взаимосвязи, присутствующие в изображениях. Результаты работы и сравнение с CFD расчетом можно наблюдать на рисунке 8, пример поля представлен на рисунке 7.

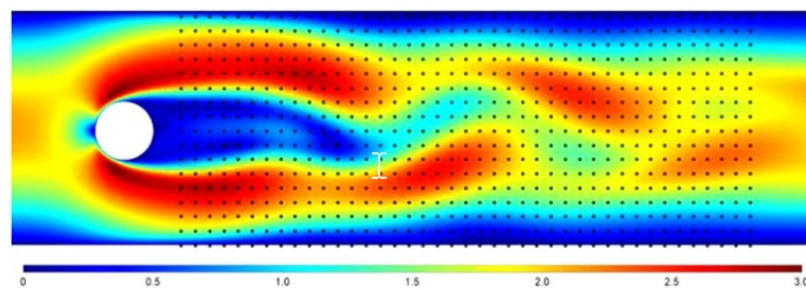


Рисунок 7. [13] Поле скорости обтекаемого цилиндра.

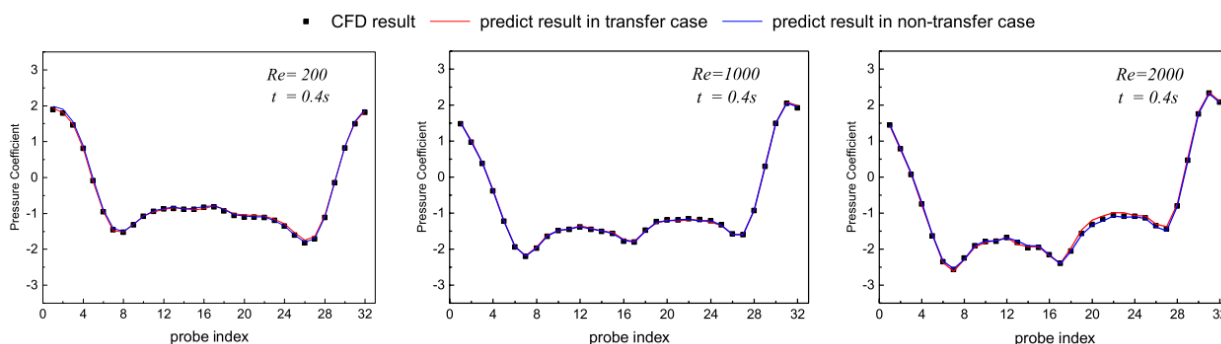


Рисунок 8. [13] Коэффициент давления в 32 точках цилиндра между предсказаниями модели и результатами CFD для различных чисел Рейнольдса в случае переноса и без переноса.

Еще одним примером применения CNN является работа [14], где сверточная нейронная сеть использовалась для прогнозирования коэффициентов сопротивления и подъемной силы на основе изображения поля скоростей. Это еще раз подтверждает эффективность CNN в задачах анализа и интерпретации пространственных данных в контексте вычислительной гидродинамики.

1.5 Dynamic (Recurrent) RNN

Динамические или рекуррентные нейронные сети (RNN) — это тип нейронных сетей, разработанный специально для работы с последовательными данными. Отличительной чертой RNN является наличие "памяти" в виде скрытых состояний, которые передают информацию из предыдущих шагов к текущему, позволяя обрабатывать информацию с учетом ее контекста во времени или последовательности. Это особенно полезно при работе с временными рядами, данных последовательностей, и при обработке данных, где контекст и последовательность важны, как например в задачах обработки естественного языка или прогнозирования временных рядов.

В работе [15] динамическая нейронная сеть обучается предсказывать температуру в 6 точках на основе данных о тепловом потоке на входе и выходе. Данные были собраны в процессе роста кристаллов с использованием 500 различных процессов и 100 шагов в каждом. Использование RNN позволило обучить модель, которая учитывает временную динамику процесса, что привело к улучшению качества прогноза. На рисунке 9 можно наблюдать

постановку задачи и пример прогнозов для различных вариантов архитектур нейронной сети.

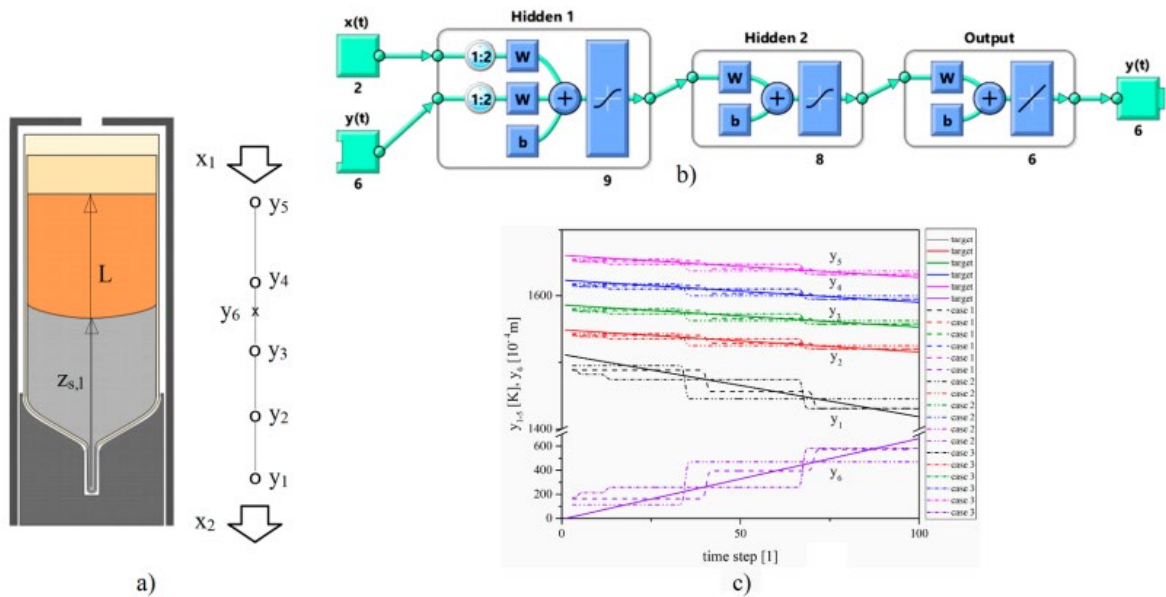


Рисунок 9. [15] Быстрое прогнозирование динамического рецепта роста при выращивании VGF-GaAs с использованием NARX-ANN: (a) конфигурация вычислительной области, (b) архитектура NARX с 2 скрытыми слоями и 2 временными задержками, (c) прогнозируемые температуры и положение интерфейса в точках мониторинга с помощью ANN.

В работе [16] использована одна из разновидностей рекуррентных нейронных сетей - LSTM (Long Short-Term Memory). Эта архитектура особенно эффективна при работе с долгосрочными зависимостями в данных. Модель обучена на основе 800 экспериментальных данных с шагом в 2 секунды. Сравнение результатов модели LSTM с другими подходами, включая SVM и DNN, показало ее высокую эффективность. На рисунке 10 можно увидеть результат прогноза диаметра кристалла в сравнении с реальными значениями.

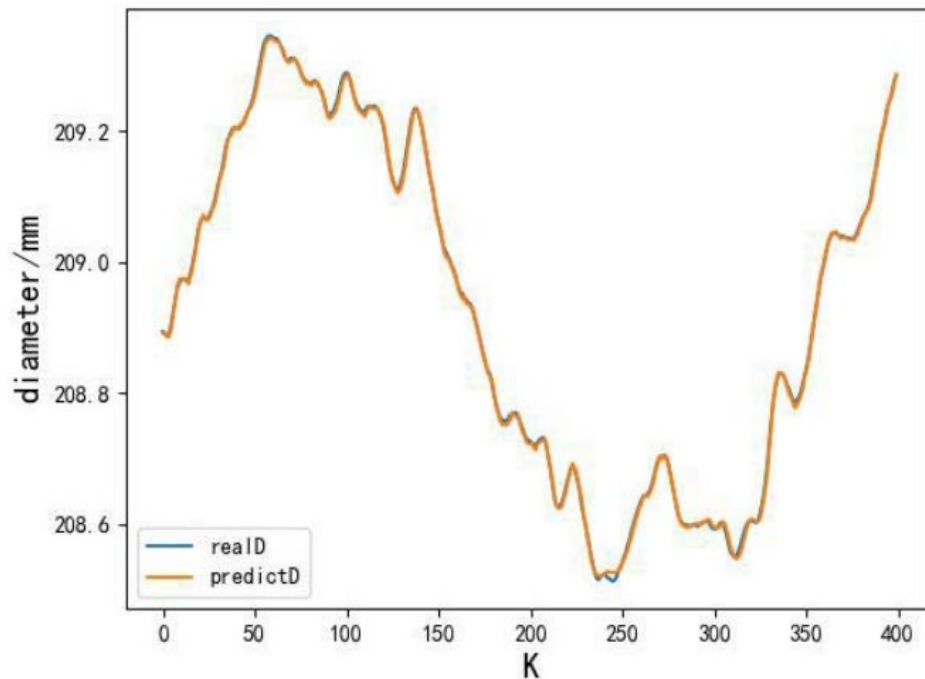


Рисунок 10. [16] Диаметр модели идентификации LSTM и фактический диаметр.

На рисунке 11 можно наблюдать сравнение значений среднеквадратичной ошибки при использовании различных методов машинного обучения.

<i>Method</i>	<i>Input vector</i>	<i>MSE</i>
<i>Support vector regression</i>	$x=(U(k-65),U(k-64), D(k-2),D(k-1))$	$1.7861E-4$
<i>BP neural network</i>	$x=(U(k-65),U(k-64), D(k-2),D(k-1))$	$5.8321E-5$
<i>LSTM identification network</i>	$x=(U(k-65),U(k-64), D(k-2),D(k-1))$	$8.9021E-6$

Рисунок 11. [16] Результаты индекса производительности.

Большинство вышеупомянутых подходов сталкивается с проблемой - большой объем данных, необходимых для обучения. Несмотря на то, что прогнозы нейронных сетей можно получить почти мгновенно, в отличие от результатов, получаемых с помощью CFD-солверов, преднамеренная генерация большого количества данных для конкретной задачи не считается целесообразной с точки зрения эффективности, временных и вычислительных затрат. В связи с этим, массовое использование этих методов маловероятно. Однако следующий подход помогает преодолеть эту проблему.

1.6 Физико-информированные нейронные сети (PINN)

PINN, или Physics-Informed Neural Networks, являются относительно новым подходом в области машинного обучения, который интегрирует физические законы в обучение нейронной сети.

В классическом машинном обучении нейронные сети обучаются на основе больших объемов данных, при этом они не имеют представления о физических законах, которые управляют процессом. Однако в многих научных и инженерных проблемах мы знаем физические законы, которые описывают систему.

PINN объединяют эти два подхода, интегрируя физические законы в форме дифференциальных уравнений в процесс обучения. Это позволяет обучать модели на гораздо меньших объемах данных, при этом достигая высокой точности и физической осмысленности прогнозов. Также часто они применяются для обратного прогноза, когда нужно спрогнозировать параметры потока на основе имеющихся данных.

В [17] PINN используются для прогноза 3D течений на основе их нескольких 2D проекций, которые можно увидеть на рисунке 12.

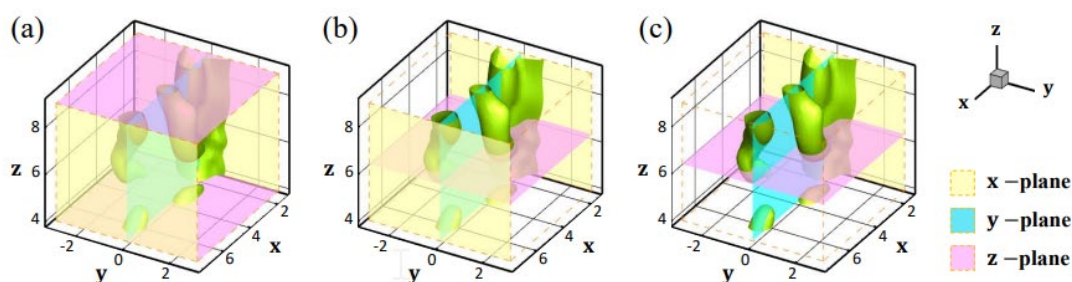


Рисунок 12. [17] Тематическое исследование PINNs для несжимаемых потоков: постановка задачи для реконструкции 3D потока по 2D2C наблюдениям. (a) Пример 1: наблюдаются две x-плоскости ($x = 1.5, 7.5$), одна y-плоскость ($y = 0$) и две z-плоскости ($z = 4.0, 9.0$). (b) Случай 2: наблюдаются две x-плоскости ($x = 1.5, 7.5$), одна y-плоскость ($y = 0$) и одна z-плоскость ($z = 6.4$). (c) Случай 3: наблюдается одна x-плоскость ($x = 1.5$), одна y-плоскость ($y = 0$) и одна z-плоскость ($z = 6.4$). Обратите внимание, что для поперечных плоскостей измеряются только проецируемые векторы. Целью является вывод трехмерного потока в исследуемой области с помощью PINNs из этих 2D2C наблюдений.

Результаты работы можно наблюдать на рисунке 13.

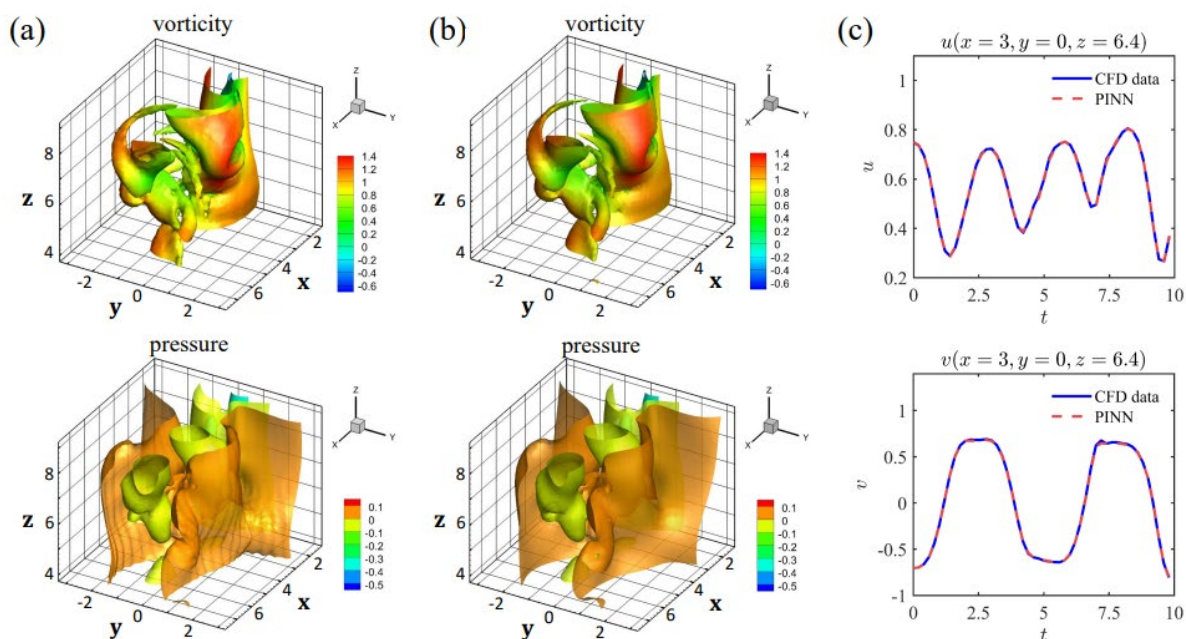


Рисунок 13. [17] Исследование PINNs для несжимаемых потоков: результат вывода PINNs для примера 2. (a) Изоповерхности величины вихря (вверху) и давления (внизу) при $t = 8,0$ по данным CFD. (b) Изоповерхности величины вихря (вверху) и давления (внизу) при $t = 8,0$, полученные с помощью PINNs. (c) Точечное измерение ($x = 3, y = 0, z = 6.4$) скорости (u, v) против времени.

В этом случае трехмерный поток определяется PINNs по четырем поперечным плоскостям.

Одним из наиболее популярных и наглядных применений PINN является исследование 3D потока над чашкой эспрессо [4, 18, 19]. В данном случае, нейронная сеть предсказывает 3D поток полей скорости и давления, исходя из 3D снимков полей температуры, полученных методом Tomo-BOS.

Постановку задачи и результаты можно наблюдать на рисунках 14, 15, 16, 17.

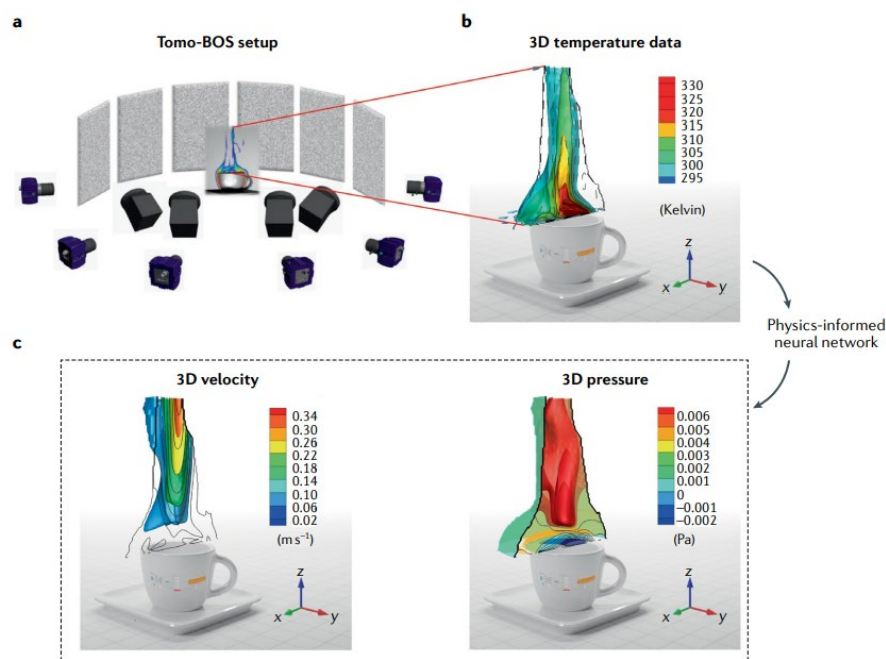


Рисунок 14. [4] Определение 3D потока над чашкой эспрессо с помощью системы визуализации Томо-BOS и нейронных сетей, основанных на физике (PiNNs). а | Шесть камер расположены вокруг чашки эспрессо, регистрируя искажение точечных узоров на панелях, расположенных на заднем плане, где искажение вызвано изменением плотности воздушного потока над чашкой эспрессо. Данные изображения получены и обработаны с помощью программного обеспечения Tomographic BOS компании LaVision (DaVis 10.1.1). б | 3D температурное поле, полученное из поля коэффициента преломления и реконструированное на основе 2D изображений со всех шести камер.

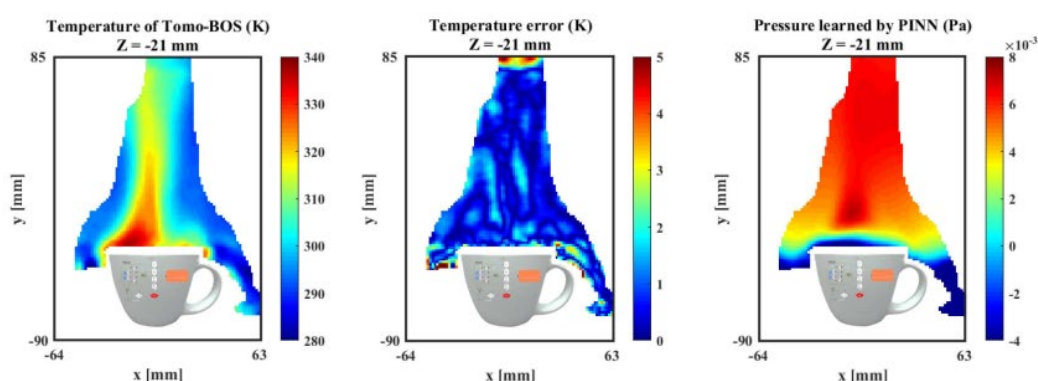


Рисунок 15. [18] Результат Томо-BOS/PINN с временной понижающей выборкой данных: Температура BOS, абсолютная ошибка температуры, предполагаемое давление при $z = -21$ мм и $t = 2,96$ с.

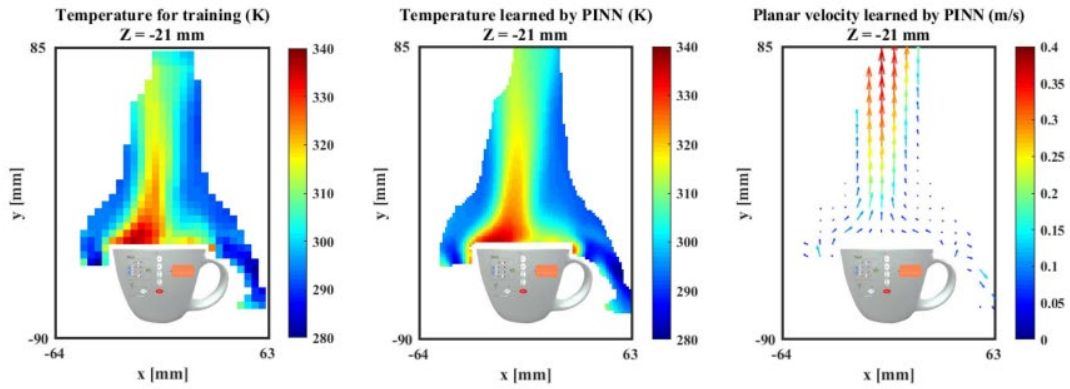


Рисунок 16. [18] Результат Томо-BOS/PINN с пространственным понижением дискретизации данных: (слева направо) пониженные температурные данные, используемые для PINN, температура, регрессированная PINN, и векторы скорости, вычисленные PINN при $z = -21$ мм и $t = 2,96$ с.

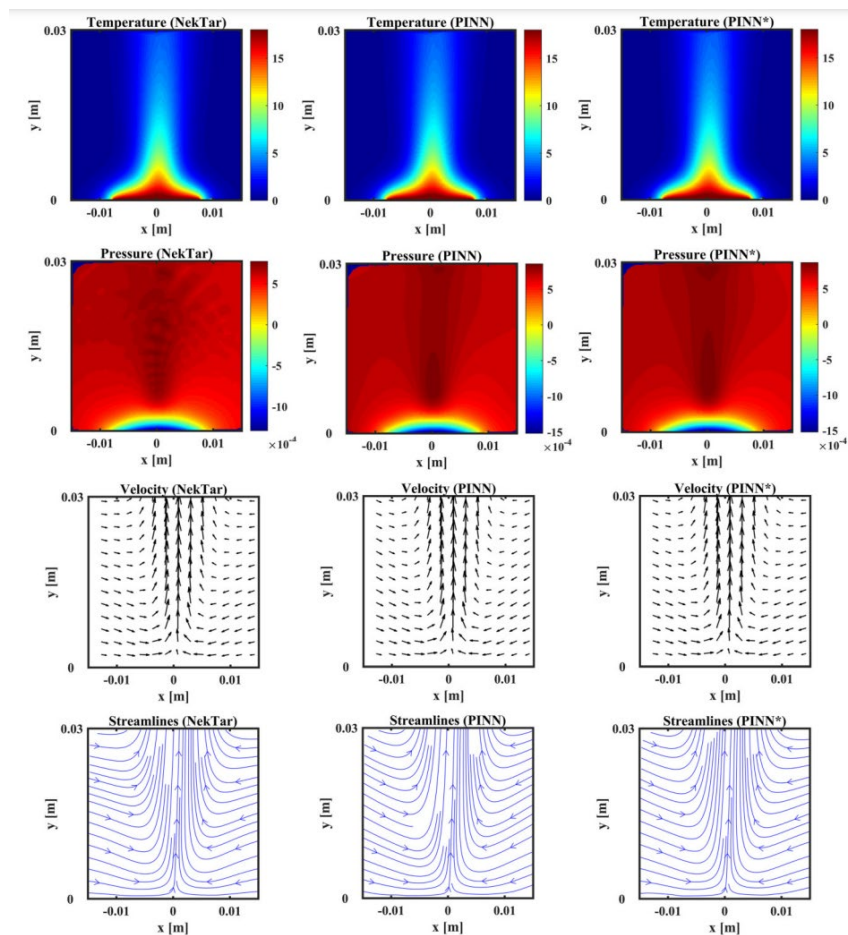


Рисунок 17. [19] Валидация с использованием синтетических данных: температура, давление, векторы скоростей и линии потоков задачи естественной конвекции при $t = 0,5$ с. Первый столбец: симуляции NekTar, второй столбец: результаты PINN с размерной установкой, третий столбец: результаты PINN с безразмерной установкой (обозначены PINN*). Ошибки L2 и RMSE размерного PINN составляют $T = 0,16\%$, $p = 3,47\%$ и $RMSE = 8,5$

$\times 10^{-4}$ м/с, а безразмерного PINN - $T = 0,10\%$, $p = 2,72\%$ и $RMSE = 6,8 \times 10^{-4}$ м/с.

Применение физически информированных нейронных сетей (Physics-Informed Neural Networks, PINNs) открывает новые возможности для достижения высокой точности прогнозирования, сохраняя при этом физическую осмысленность результатов. Это обеспечивается за счет интеграции знаний о физических законах прямо в структуру модели, что позволяет добиться качественных результатов даже при ограниченном объеме обучающих данных.

Благодаря сочетанию преимуществ PINNs и других подходов к обучению нейронных сетей, возможно эффективное моделирование сложных гидродинамических систем, основанное на существующих физических законах.

Такие преимущества как возможность эффективного применения при ограниченном объеме данных, высокая точность прогнозов, а также сохранение физической осмысленности решений делают применение PINNs особенно актуальным в контексте гидродинамики. Это определяет важность и актуальность изучения данного подхода, который стал ключевым в рамках нашей работы.

2. ПОСТАНОВКА ЗАДАЧИ, МАТЕМАТИЧЕСКАЯ МОДЕЛЬ И ПРЕДВАРИТЕЛЬНЫЕ РАСЧЕТЫ ДЛЯ ЗАДАЧИ ТЕЧЕНИЯ В КАНАЛЕ С ВНЕЗАПНЫМ СУЖЕНИЕМ

2.1 Математическая модель

Нами рассматривалась задача прогноза полей скорости и давления для задач стационарного ламинарного течения в канале с внезапным сужением. На рисунке 18 представлена схема течения.

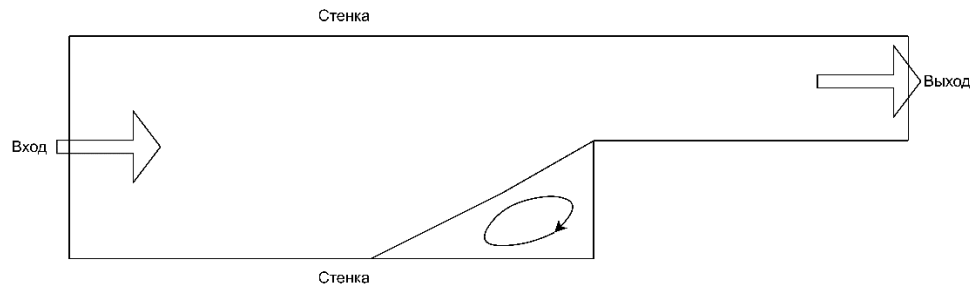


Рисунок 18. Схема течения.

Такое течение описывается стационарными уравнениями Навье-Стокса:

$$\begin{cases} \nabla \cdot \vec{v} = 0 \\ (\vec{v} \cdot \nabla) \vec{v} = -\frac{1}{\rho} \nabla p + \frac{\mu}{\rho} \nabla^2 \vec{v} \end{cases} \quad (1)$$

где

$$\vec{v} = (u, v) \quad (2)$$

В стандартной формулировке эти уравнения включают производные второго порядка, что приводит к определенным сложностям при применении физико-информированных нейронных сетей (PINN), предпочитающих работу с первыми производными. В этой связи, мы переформулировали уравнения, заменив вторые производные на первые, используя тензор напряжений:

$$\begin{cases} \nabla \cdot \vec{v} = 0 \\ (\vec{v} \cdot \nabla) \vec{v} = -\frac{1}{\rho} \nabla \sigma + \vec{g} \end{cases} \quad (3)$$

где σ – тензор напряжений:

$$\sigma = p\mathbf{E} + \mu(\nabla \vec{v} + \nabla \vec{v}^T) \quad (4)$$

где \mathbf{E} – единичный тензор.

В инструментарии, используемом для работы с физико-информированными нейронными сетями, мы отходим от прямого расчета скорости. Вместо этого, мы используем функцию тока, что позволяет нам более эффективно работать с уравнением неразрывности. Таким образом, скорость вычисляется из функции потока по следующему уравнению:

$$[u, v, 0] = \nabla \times [0, 0, \psi] \quad (5)$$

Для двумерной задачи тензор напряжений можно написать следующим образом:

$$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \quad (6)$$

где

$$\sigma_{11} = -p + 2\mu \frac{\partial u}{\partial x} \quad (7)$$

$$\sigma_{22} = -p + 2\mu \frac{\partial v}{\partial y} \quad (8)$$

$$\sigma_{12} = \sigma_{21} = \mu \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \quad (9)$$

Отсюда:

$$p = -\frac{\sigma_{11} + \sigma_{22}}{2} = -\frac{tr(\sigma)}{2} \quad (10)$$

Используемые в рассматриваемой задаче граничные условия определяются следующим образом:

На входе задается профиль по формуле:

$$u = 4 * \frac{u_{max} y (y_{max} - y)}{(y_{max})^2} \quad (11)$$

На выходе давление равно нулю

$$p(x_{outlet}, y_{outlet}) = 0 \quad (12)$$

На стенках – условие прилипания:

$$v(x_{wall}, y_{wall}) = 0, u(x_{wall}, y_{wall}) = 0 \quad (13)$$

2.2 Выбор расчетной области и сетки для проведения эталонных расчетов в пакете Fluent

Эталонные расчеты с применением пакета ANSYS Fluent проводятся в целях создания репрезентативного набора данных, который затем используется для сравнительного анализа результатов, генерируемых обученной нейронной сетью. Результаты, полученные при помощи Fluent, считаются эталонными для большого спектра задач. Использование данного пакета, таким образом, позволяет гарантировать, что процесс валидации нейронной сети основывается на данных, наиболее точно отражающих действительные физические процессы.

Использовалась ламинарная модель расчета, число Рейнольдса, как и в случае с нейросетью варьировалось за счет изменения вязкости, условия на входе задавались с помощью выражения, согласно формуле (11). Использованная схема расчета – Coupled для решения уравнений движения и уравнения непрерывности одновременно. Для вычисления градиентов использовался метод наименьших квадратов (Least Squares Cell Based). Также использовалась дискретизация второго порядка для уравнений давления (Second Order) и импульса (Second Order Upwind).

В контексте рассматриваемой задачи, основной фокус нашего анализа направлен на область течения в непосредственной близости от зоны рециркуляции, которая формируется в районе места сужения. Основное внимание уделяется исследованию области течения, в которой проявляются ключевые зависимости от числа Рейнольдса (Re) и высоты сужения (h).

Чтобы провести более точное исследование, важно исключить влияние других параметров, например, длины участка перед сужением. В этом контексте, преимущественное значение имеет исключение влияния входной границы на характеристики течения в области рециркуляции, что иллюстрируется на рисунке 19.

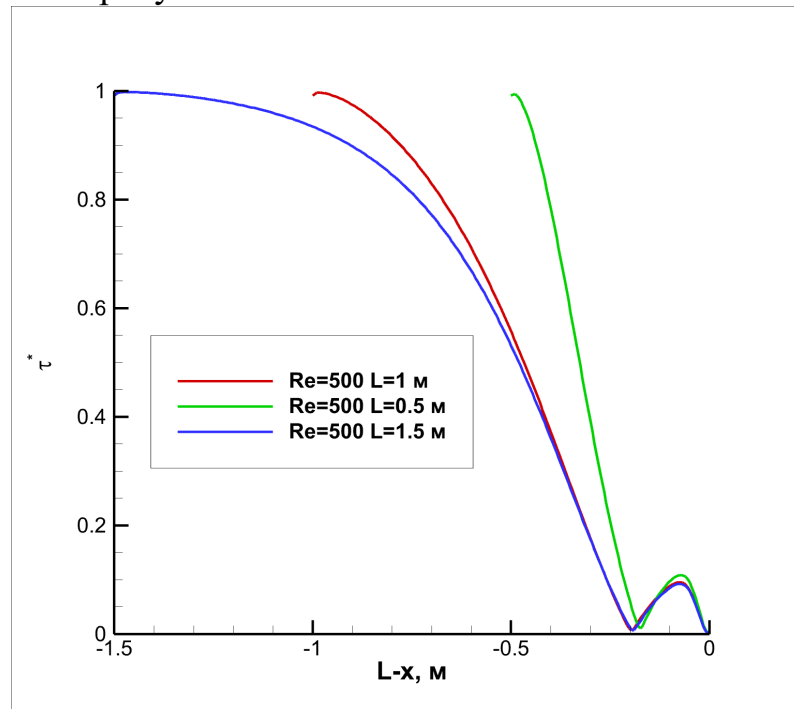


Рисунок 19. График безразмерного напряжения трения на нижней стенке в зависимости от длины канала при числе Рейнольдса Re=500.

Здесь используется τ^* - безразмерный параметр напряжений трения на нижней стенке, рассчитывается по следующей формуле (15), вывод которой описан ниже.

$$\tau = -\mu \frac{\partial u}{\partial y} \quad (14)$$

Учитывая, что на входе профиль задается формулой (11). Мы получаем следующие уравнения:

$$\begin{aligned} \tau_{inlet} &= 4\mu \frac{u_{max}y}{y_{max}^2} + 4\mu \frac{u_{max}(y_{max} - y)}{y_{max}^2} \\ \tau_{inlet}(y = 0) &= 4\mu \frac{u_{max}}{y_{max}} \end{aligned}$$

Откуда:

$$\tau^* = \frac{\tau_w y_{max}}{4\mu u_{max}} \quad (15)$$

При рассмотрении длины канала перед сужением L и длины зоны рециркуляции, мы получаем следующие результаты:

Таблица 1.

Длина канала перед сужением L , м	Длина зоны рециркуляции, м	Разница с предыдущим, %
0.5	0.1741	0
1	0.1935	11
1.5	0.1933	0.089%

На основании полученных результатов, было решено выбрать длину канала перед сужением равной $L = 1$ м. Для участка после расширения длина $l = 0.5$ м. Высота канала составляет $H = 0.41$ м, а высота сужения $h = 0.21$ м.

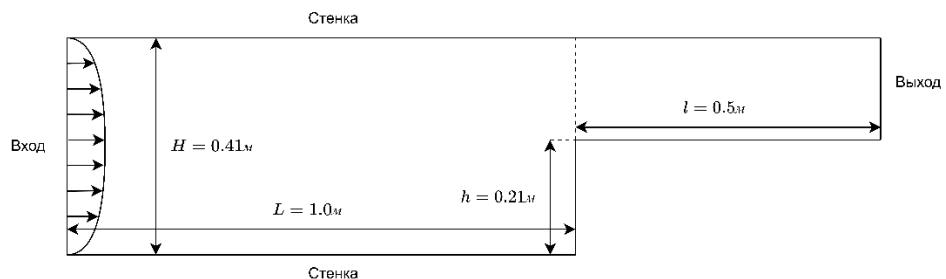


Рисунок 20. Схема расчетной области с указанными размерами.

Также было проведено исследование сеточной сходимости задачи, результаты которого можно наблюдать в таблице 2 и на рисунке 21.

Таблица 2.

Размер ячейки сетки, м	Длина зоны рециркуляции, м	Разница с предыдущим, %
0.04	0.1639	0
0.02	0.1904	14
0.01	0.1923	1

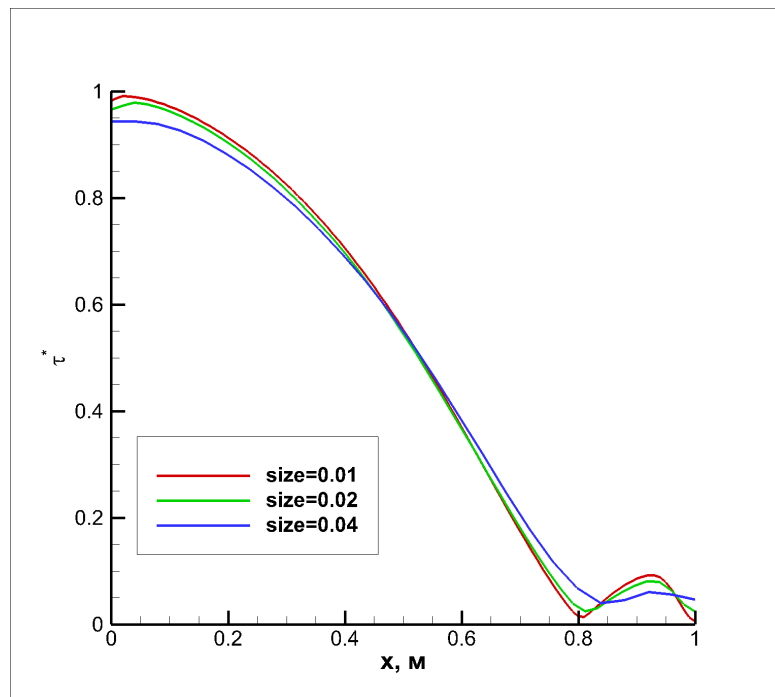


Рисунок 21. График безразмерного напряжения трения на нижней стенке для различных сеток Fluent.

В результате исследования сеточной сходимости было принято решение выбрать сетку размером ячейки в 0.01. Пример такой сетки представлен на рисунке 22.

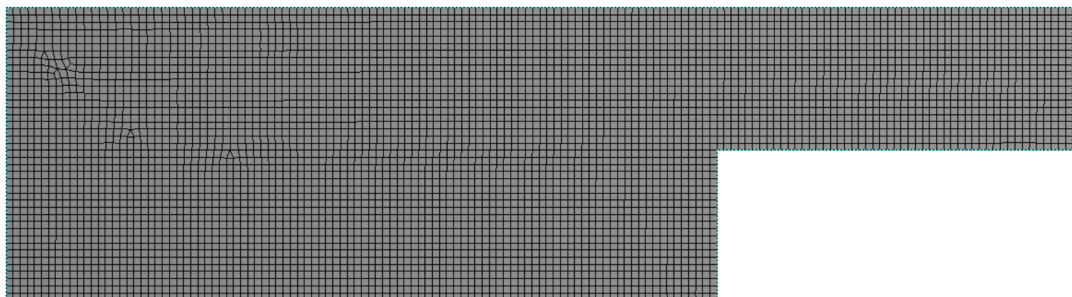


Рисунок 22. Сетка Fluent.

3. УСТРОЙСТВО PINN И ИСПОЛЬЗУЕМЫЕ ИНСТРУМЕНТЫ

3.1 Архитектура и принцип работы физико-информированной нейронной сети

В рамках исследования была разработана программа нейронной сети на языке программирования Python. Основу для создания архитектуры и проверки функционала программы обеспечивал код, представленный в [3]. Этот код первоначально разрабатывался для решения задачи стационарного ламинарного обтекания цилиндра.

В отличие от исходного кода, представленного в [3], разработанный в данной работе код был реализован с использованием Tensorflow версии 2. Это обеспечило более простой и быстрый процесс обучения на виртуальных машинах и позволило интегрировать в код новые, более эффективные методы обучения, что в конечном итоге упростило процесс оптимизации.

Архитектура нейронной сети представлена на рисунке 23.

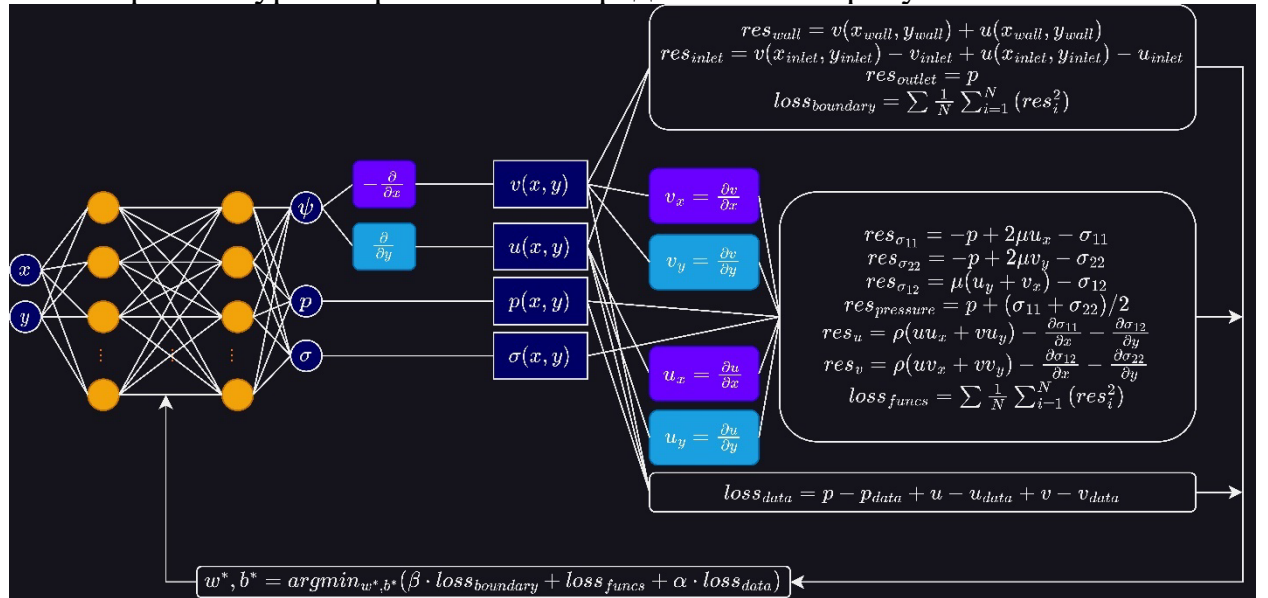


Рисунок 23. Архитектура нейронной сети.

Согласно представленной на рисунке архитектуре, нейронная сеть принимает на вход тензоры x и y размерностью $(N, 1)$, объединяет их в тензор $(N, 2)$ и предсказывает пять величин в виде тензоров $(N^2, 1)$ – функцию тока ψ , давление p , и тензор напряжений σ , а точнее три его компоненты: σ_{11} , σ_{22} , σ_{12} .

Обучение нейронной сети осуществляется с использованием метода градиентного спуска, который является итеративным методом оптимизации для поиска минимума функции. Нейронная сеть обучается, минимизируя так называемую функцию потерь, которая в нашем случае выражается как:

$$loss = \beta \cdot loss_{boundary} + loss_{funcs} + \alpha \cdot loss_{data}$$

где

- $loss_{boundary}$ - часть функции потерь, связанная с выполнением граничных условий
- $loss_{funcs}$ - часть функции потерь, связанная с выполнением уравнений Навье-Стокса
- $loss_{data}$ - часть функции потерь, рассчитанная при наличии обучающих данных.

Важно отметить, что в рамках нашего исследования компонент $loss_{data}$ не использовался.

Градиент функции потерь в отношении весов (w^*) и смещений (b^* - bias) модели указывает в направлении наибольшего увеличения этой ошибки. Таким образом, для минимизации ошибки веса и смещения нейронов обновляются в направлении, противоположном градиенту. Отклонения, или смещения, представляют собой дополнительный параметр в модели нейронной сети, который позволяет сдвигать функцию активации. Пример устройства нейрона с использованием смещений, весов и функции активации можно увидеть на рисунке 24.

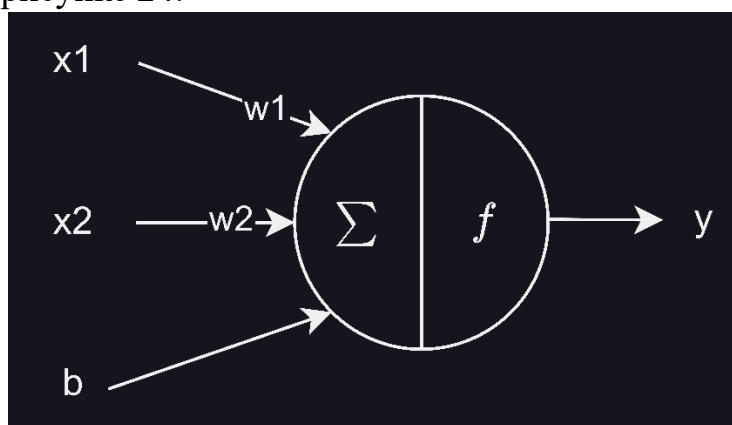


Рисунок 24. Схема одного нейрона нейронной сети.

Процесс дифференцирования и получения градиентов выполняется с применением функции GradientTape из библиотеки Tensorflow, используемой для автоматического дифференцирования. Автоматическое дифференцирование представляет собой методологический подход к вычислению производных, основываясь на последовательном разбиении функции на ряд элементарных операций (например, сложение, умножение, математические функции, такие как \sin и \exp и т.д.) и пошаговом применении правил дифференцирования к этим операциям в соответствии с правилами дифференциального исчисления.

В области расчета производных можно выделить три основных подхода: численное дифференцирование, символьное дифференцирование и автоматическое дифференцирование.

Численное дифференцирование базируется на методе конечных разностей. Однако при применении данного метода могут возникнуть

проблемы с точностью и стабильностью вычислений, особенно при работе со сложными функциями или функциями, имеющими большие производные.

Символьное дифференцирование представляет собой автоматизированный подход к формированию формул для вычисления производной функции. Несмотря на то, что данный подход обеспечивает точные результаты, он может оказаться более трудоемким и вычислительно затратным, особенно при работе со сложными функциями.

Автоматическое дифференцирование устраняет многие из этих проблем. Оно обеспечивает вычислительную эффективность и точность, позволяя вычислять производные сложных функций с высокой точностью и разумными вычислительными затратами. Это особенно важно в области машинного обучения, где требуется вычислять градиенты сложных функций потерь относительно большого количества параметров.

Для расчета значений $loss_{boundary}$ и $loss_{funcs}$ предсказанные нейронной сетью параметры подставляются в соответствующие уравнения и граничные условия. В результате этих операций генерируется тензор размерности $(N^2, 1)$ для каждого уравнения, содержащий невязки, получаемые в процессе подстановки. Следующим шагом является извлечение среднеквадратичного значения из этого тензора, которое затем включается в функцию ошибок.

В нейронную сеть также можно загрузить дополнительные данные для усовершенствования процесса обучения. В процессе суммирования ошибок, полученных из уравнений, граничных условий и данных, используются коэффициенты β и α , заранее задаваемые пользователем. Эти коэффициенты позволяют регулировать значимость каждого элемента в общей ошибке. В данной работе осуществлялся подбор параметра β , результаты которого приведены ниже. Однако коэффициент α не применялся, поскольку как было указано выше – при обучении нейронной сети данные не были задействованы.

3.2 Тестирование разработанной программы

Для проверки написанного кода была взята аналогичная [3] задача обтекания цилиндра, с геометрией показанной на рисунке 25.

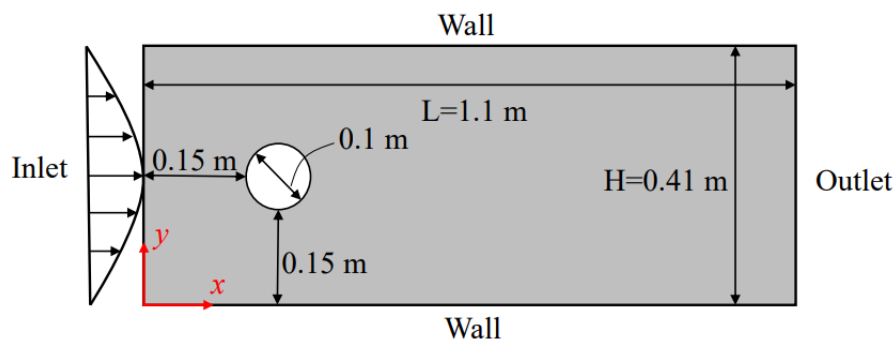


Рисунок 25. [3] Схема расчетной области.

Входной профиль задавался в соответствии с формулой (11), на стенках было задано условие прилипания (13), а на выходе установлено условие равенства нулю давления (12).

Рассмотренная задача была реализована в программе, описанной в [3] (Tensorflow 1), Fluent, а также в программе, разработанной в рамках данной работы (Tensorflow 2).

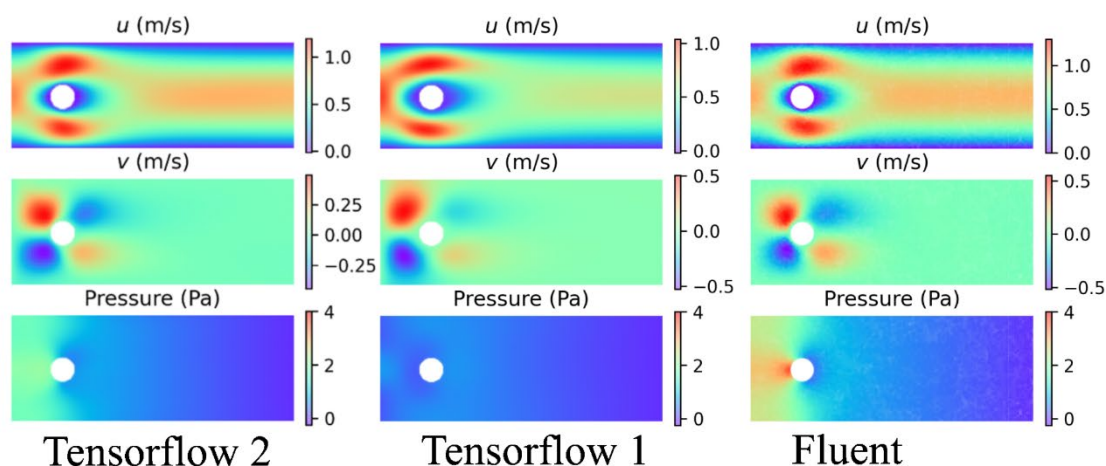


Рисунок 26. Проверка модифицированного кода - результаты, получаемые после 10 000 итераций.

Изучение представленного рисунка позволяет заметить, что после проведения 10 000 итераций наблюдается формирование поля продольной компоненты скорости (u). В частности, зафиксированы высокие значения этой величины в более узких сечениях, а также замечен скоростной профиль, формирующийся в зоне за цилиндром.

Что касается сформировавшейся поперечной компоненты скорости (v), ее анализ также указывает на физичность полученных результатов, что подтверждается наличием положительной и отрицательной компонент при обтекании цилиндра потоком.

Поле давления к 10 000 итерациям еще не успевает полностью сформироваться, однако уже прослеживаются начальные стадии его формирования, которые проявляются в увеличении значений в начале канала.

Таким образом, анализ полученных полей демонстрирует успешность реализации кода, разработанного в рамках данного исследования.

4. РЕЗУЛЬТАТЫ РАБОТЫ PINN И ИХ ОБСУЖДЕНИЕ

4.1 Описание процесса обучения сети

В ходе формирования нейронной сети происходит определение основных параметров обучения, включая функцию активации, алгоритм инициализации, количество слоев и количество нейронов в каждом слое. После этого метод выборки латинского гиперкуба (Latin Hypercube Sampling) применяется для создания равномерно распределенных координат в области канала и на его границах. Таким образом, образуется набор точек, равномерно распределенных в пространстве, не формируя, тем не менее, сетки в традиционном понимании этого понятия, такую «сетку» часто называют стохастической. Данные точки затем используются в процессе обучения нейронной сети, которая стремится приблизить решение уравнений в заданных точках.

Процедура обучения выполняется следующим образом: инициализатор используется для начальной инициализации весов с элементом случайности, а также для инициализации смещений. Далее выполняется первый прогноз сети, в ходе которого вычисляется функция потерь. На основе результатов вычисления функции потерь нейронная сеть корректирует веса и смещения.

Рисунок 27 ниже демонстрирует типичный график процесса обучения:

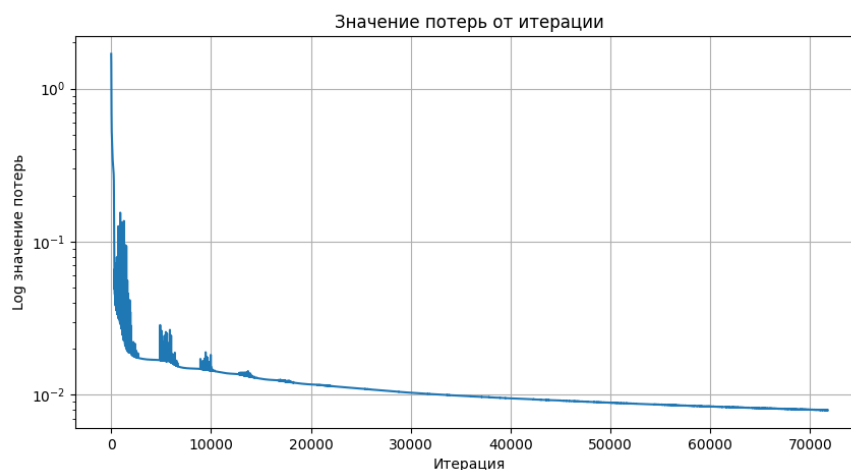


Рисунок 27. График значения функции ошибок в зависимости от итерации обучения PINN.

На графике отчетливо видно резкое падение функции ошибок на начальных стадиях обучения. Встречаются всплески, связанные с особенностями изменения шага обучения (learning rate). По мере продвижения по итерациям скорость снижения ошибки значительно уменьшается.

Следует отметить, что обучение физико-информированной нейронной сети (PINN) является довольно трудоемким процессом по времени. На видеокарте Tesla T4 16 GB и процессоре Intel(R) Xeon(R) CPU @ 2.00GHz, обучение сети на протяжении 10 000 итераций занимает приблизительно 35 минут. В среднем для обучения потребовалось 80 000 итераций. Однако, после завершения обучения нейросеть способна выдавать результаты в пределах нескольких секунд. Эта способность быстрого прогнозирования может быть крайне полезна в ряде практических ситуаций, где требуется быстрое и точное решение задачи.

4.2 Подбор гиперпараметров

В контексте нейронных сетей гиперпараметры — это конфигурационные параметры, которые задаются до начала процесса обучения и остаются фиксированными на протяжении всего этого процесса. Они не подлежат автоматической коррекции, в отличие от весовых коэффициентов модели, которые обновляются на каждой итерации обучения. В данной работе в качестве гиперпараметров рассматривались количество скрытых слоев и количество нейронов в каждом из этих слоев, типы используемых оптимизаторов и функций активации, размер шага обучения (learning rate), а также тип функции потерь.

Отбор гиперпараметров проводился путем варьирования и анализа двух ключевых показателей: скорости обучения, оцениваемой через динамику снижения значения функции потерь (значения потерь), а также времени, требуемого для обучения. Стоит отметить, что последний показатель проявлял существенные различия в основном при изменении количества скрытых слоев. Понятно, что чем ниже показатели скорости обучения и значения функции потерь, тем эффективнее выбранный набор гиперпараметров.

4.2.1 Выбор функции активации

Функция активации в нейронной сети представляет собой математическую операцию, применяемую к выходному значению каждого нейрона.

Существует множество различных типов функций активации, каждая из которых имеет свои преимущества и недостатки. Например, линейная функция активации проста для вычисления, но может быть использована только для простых задач. Сигмоидная функция активации преобразует входное значение в диапазон от 0 до 1, что может быть полезно для вероятностных предсказаний.

Для улучшения точности прогнозов в данной работе было проведено сравнение различных функций активации.

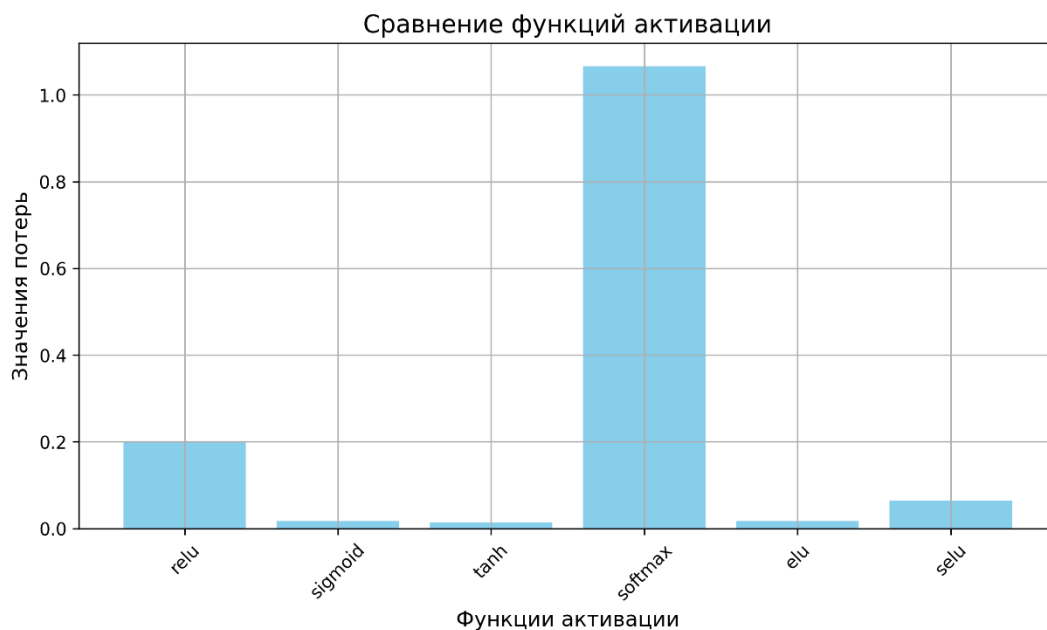


Рисунок 28. Диаграмма потерь нейронной сети после 5 000 итераций, в зависимости от выбора функции активации.

По итогам проведенных экспериментов в качестве оптимальной была выбрана функция активации \tanh (согласно [20], данная функция также демонстрировала высокую эффективность при использовании в PINNs).

4.2.2 Выбор инициализатора

Инициализаторы в контексте нейронных сетей относятся к способу начального задания значений весовых коэффициентов сети перед началом процесса обучения. Правильная инициализация весов может иметь значительное влияние на скорость и эффективность обучения нейросети.

Для функции активации \tanh рекомендуется использовать инициализацию Глорота (glorot normal или glorot uniform), также известную как инициализация Xavier. Этот метод инициализации разрабатывался специально для функций активации, симметричных относительно нуля. Он также рекомендуется в качестве метода инициализации для PINN в [21].

Кроме того, были проведены эксперименты с инициализацией lescun normal и lescun uniform, предназначенными для нейронных сетей с сигмоидной функцией активации, но потенциально применимыми и к сетям с функцией активации \tanh .

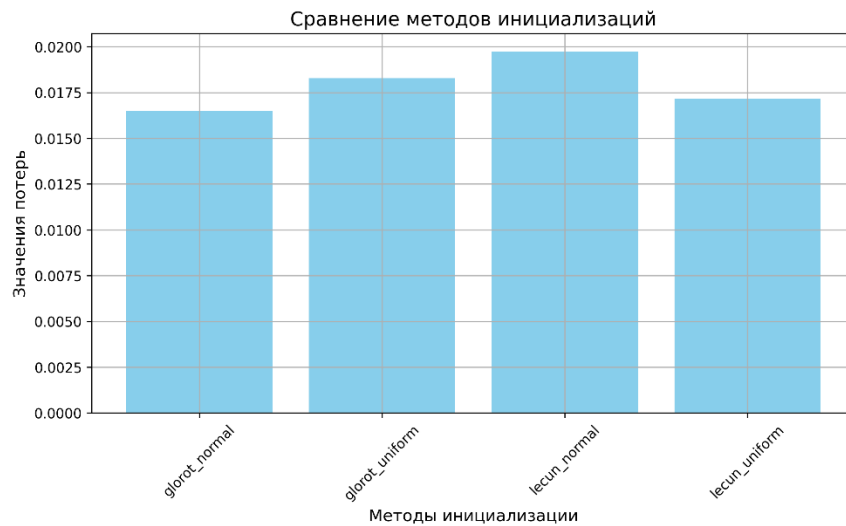


Рисунок 29. Диаграмма потерь нейронной сети после 5 000 итераций, в зависимости от выбора методов инициализации.

На основе полученных диаграмм было принято решение выбрать инициализацию gloriot uniform.

4.2.3 Выбор оптимального числа нейронов и скрытых слоев

Одним из ключевых аспектов в конфигурации нейронных сетей является определение подходящего количества нейронов и скрытых слоев — параметров, которые могут значительно повлиять на обучаемость и предсказательные способности модели.

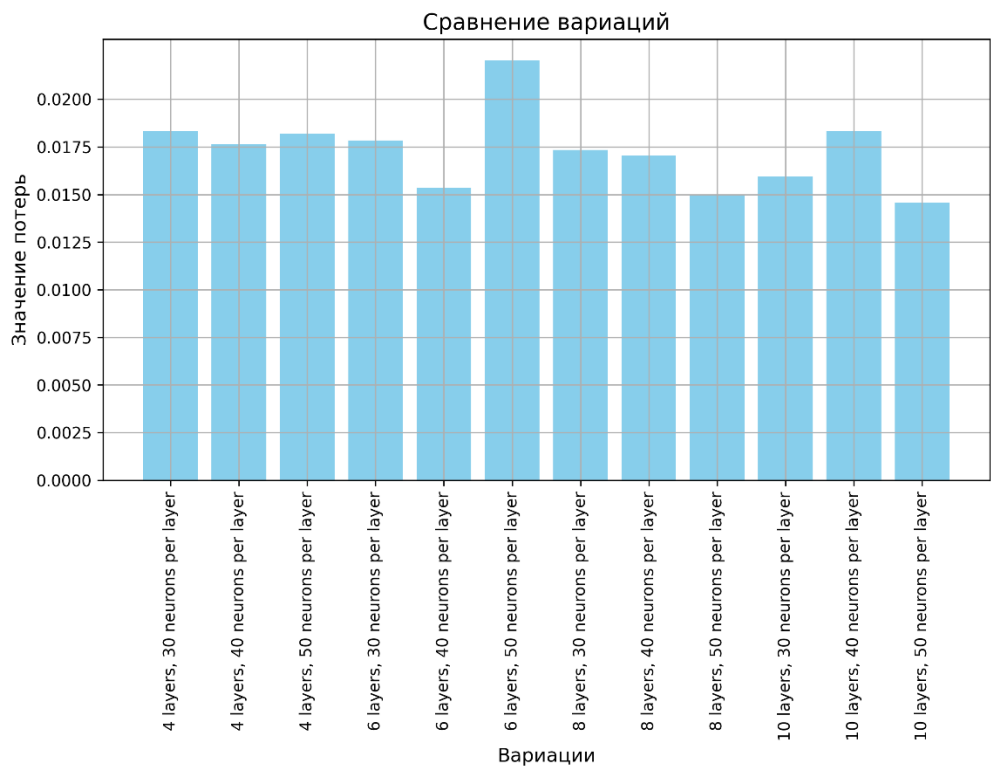


Рисунок 30. Диаграмма потерь нейронной сети, после 5 000 итераций обучения, в зависимости от числа слоев и нейронов на каждом слое.

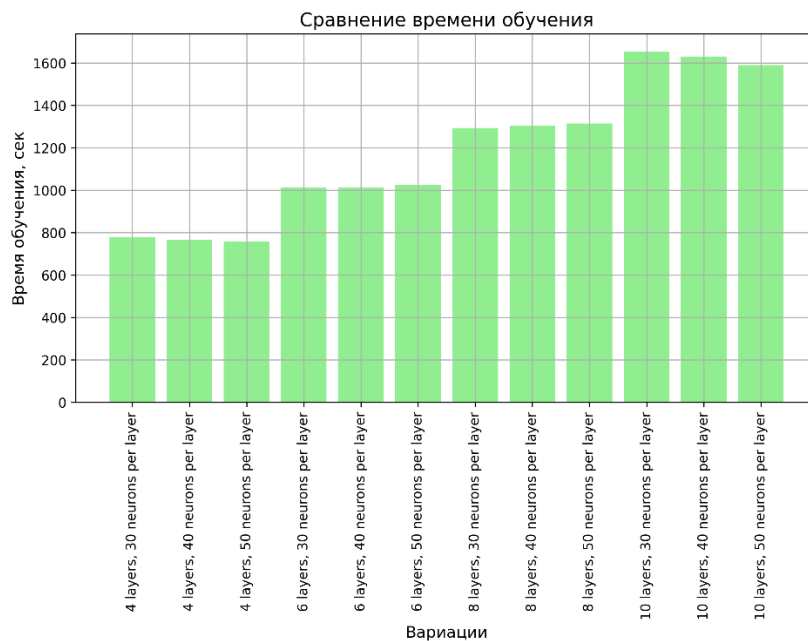


Рисунок 31. Диаграмма, показывающая время обучения 5 000 итераций нейронной сети в зависимости от числа слоев и нейронов на каждом слое.

Опираясь на результаты проведенного анализа и визуализированных на графиках данных, было принято решение о конфигурации нейронной сети, состоящей из 6 слоев с 40 нейронами в каждом. Это позволяет достичь

оптимального соотношения между качеством обучения и временем, затраченным на это обучение.

4.2.4 Выбор оптимизатора

Выбор подходящего оптимизатора для обучения нейронных сетей является еще одним значимым этапом в процессе конфигурации и может существенно повлиять на качество обучения и эффективность предсказаний.

SGD (Stochastic Gradient Descent): базовый метод оптимизации, который использует стохастический градиентный спуск для обновления весов модели.

RMSprop (Root Mean Square Propagation): оптимизатор, который использует адаптивный выбор скорости обучения, чтобы улучшить стабильность и производительность обучения.

Adam (Adaptive Moment Estimation): один из самых популярных методов оптимизации, который сочетает в себе преимущества методов RMSprop и SGD с импульсом.

Adamax: вариация оптимизатора Adam, предлагающая более стабильную оценку адаптивных скоростей обучения.

Nadam: расширение оптимизатора Adam, которое включает в себя стратегию Nesterov Accelerated Gradient, улучшая эффективность обучения.

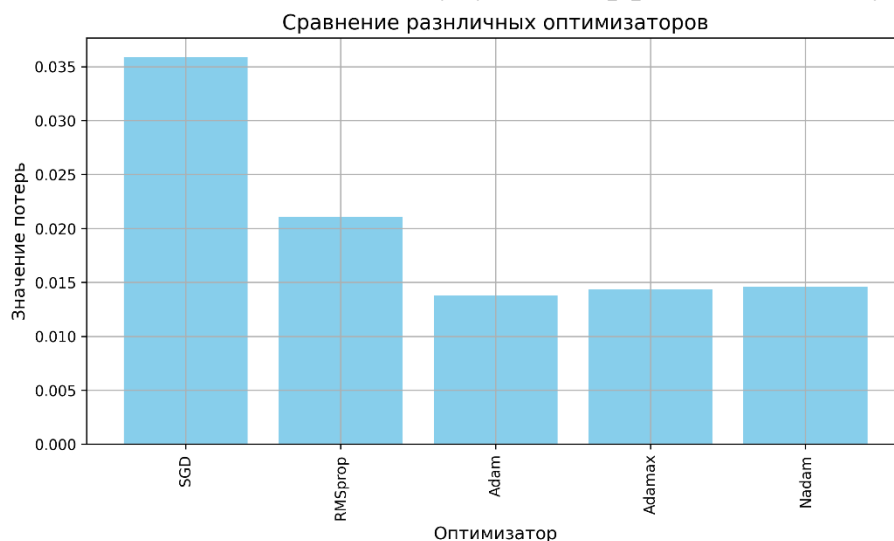


Рисунок 32. Диаграмма потерь нейронной сети после 10 000 итераций, в зависимости от выбора оптимизатора.

На основании анализа результатов обучения с использованием различных оптимизаторов было принято решение о применении оптимизатора Adam, показавшего наиболее эффективные результаты в данной задаче.

4.2.5 Выбор коэффициента β при граничных условиях

Коэффициент β представляет собой параметр, регулирующий вес потерь, возникающих при нарушении граничных условий в процессе

обучения. Этот коэффициент играет важную роль, указывая оптимизатору на степень важности учета данных условий при формировании прогнозов.

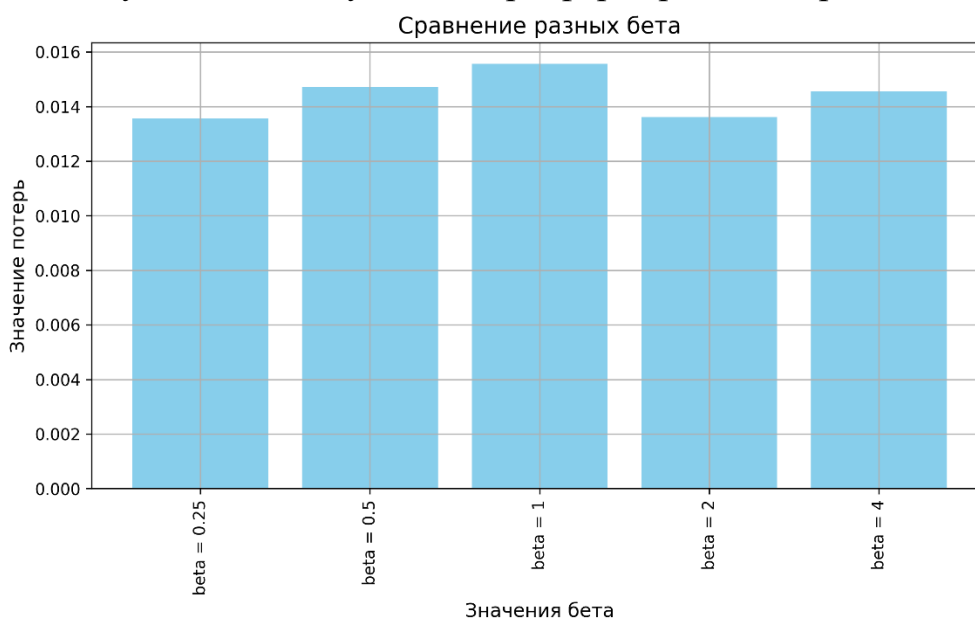


Рисунок 33. Диаграмма потерь нейронной сети, после 5 000 итераций обучения, в зависимости от коэффициента β .

В соответствии с полученными данными, было принято решение использовать в модели коэффициент β , равный двум.

4.2.6 Выбор сетки обучения

В ходе выполнения исследования была рассмотрена зависимость результатов обучения нейросети от количества точек, генерируемых с помощью метода латинских гиперкубов (LHS). Была предпринята попытка определить, как изменение числа точек влияет на качество модели после определенного числа итераций обучения.

Базовое количество точек было установлено исходя из рекомендаций, представленных в работе [3]. После этого была проведена серия экспериментов, в ходе которых базовое количество точек изменялось путем применения коэффициента вариации.

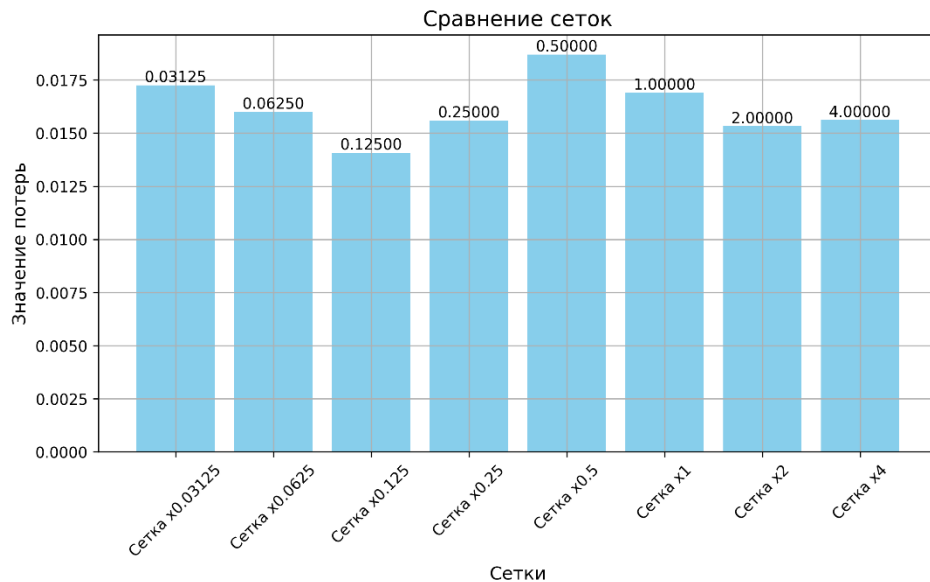


Рисунок 34. Диаграмма потерь нейронной сети, после 5 000 итераций обучения, в зависимости от заданного коэффициента стохастической «сетки».

Анализ полученных данных позволил определить оптимальное значение коэффициента для формирования "сетки" из точек. В результате было принято решение использовать значение коэффициента 0.125, что в 8 раз уменьшает изначальное количество точек.

4.3 Сравнение результатов работы PINN и Fluent

В данной работе были выполнены расчеты для различных чисел Рейнольдса, а полученные результаты были сопоставлены с данными, полученными при помощи ANSYS Fluent.

Нейронная сеть, после 80 000 итераций обучения для числа Рейнольдса $Re \approx 14$ выдает следующие результаты:

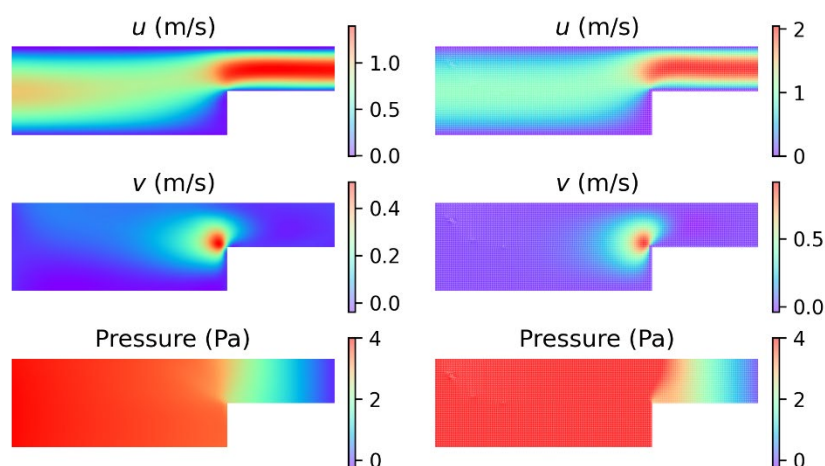


Рисунок 35. Расчет на 80 000 итераций слева - PINN, справа - Fluent для $Re \approx 14$.

Анализ полученных полей позволяет судить об успешном обучении сети.

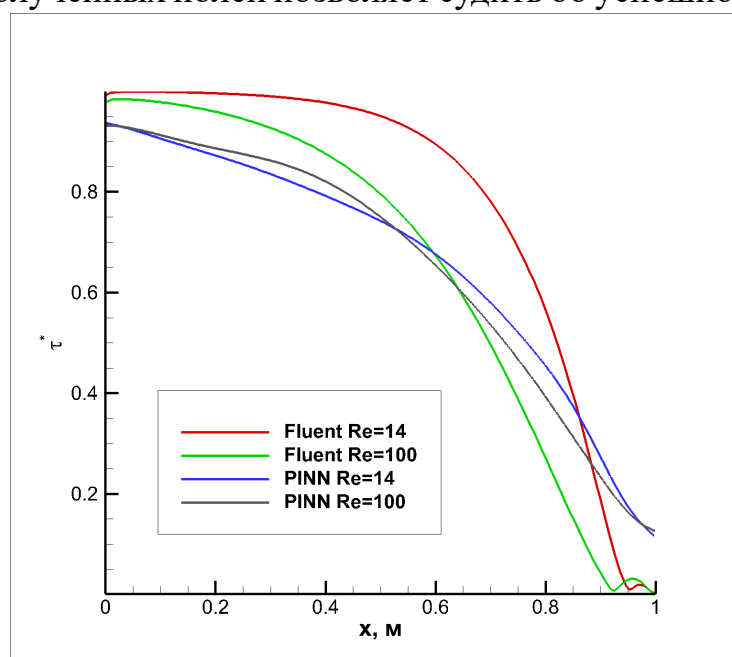


Рисунок 36. График безразмерного напряжения трения на нижней стенке при $Re = 14$ и $Re=100$ и 80 000 итераций.

На рисунке 36 можно наблюдать значительные различия в графиках безразмерного напряжения трения с максимальным отклонением значений на уровне 22%. Значение безразмерного напряжения начинается в точке со значением 0.93, а не 1, как предполагалось, и заканчивается на значении 0.11, а не 0.

В таблице 3 представлены значения процентного отклонения от данных, полученных в Fluent, после 80 000 итераций обучения физико-

информированной нейронной сети (PINN). Для расчета процентного отклонения применяется следующая формула:

$$u_{\%} = 100\% \frac{\text{mean}(|u - u_{data}|)}{\max(u_{data})}$$

Таблица 3.

Re	$u_{\%}, \%$	$v_{\%}, \%$	$p_{\%}, \%$	$\tau_{\%}^*, \%$
14	24.5	9.6	43.9	22
100	24.7	7.8	53.3	19

Анализ представленных в таблице данных позволяет увидеть, что отклонение компоненты продольной скорости (u) от эталонных значений, полученных с использованием Fluent, в обоих случаях составляет около 24%. Это отклонение можно считать довольно значительным, и стоит ожидать, что увеличение числа итераций может способствовать его сокращению.

Отклонение компоненты поперечной скорости (v) от эталонных значений немного меньше и составляет около 9% и 8% для чисел Рейнольдса 14 и 100 соответственно. Это может указывать на то, что модель с большей точностью смогла предсказать поперечные движения в потоке.

Существенные различия наблюдаются в отклонениях значений давления, которые достигают 44% и 53% для чисел Рейнольдса 14 и 100 соответственно. Это значительное отклонение может быть исследовано в будущих работах.

Ошибки в значениях напряжения трения составляют 22% и 19% для чисел Рейнольдса 14 и 100 соответственно, что также является достаточно высокими значениями.

Следует заметить, что хотя отклонения и достигают довольно высоких значений, в целом тенденции изменения физических величин в пространстве, показанные на рисунке 35, предсказываются нейронной сетью корректно.

При высоких значениях числа Рейнольдса (более 100) было замечено значительное ускорение снижения ошибки нейронной сети и вывод абсолютно нефизичных результатов, что может указывать на попадание алгоритма в локальный минимум. Пример такого вывода можно увидеть на рисунках 37 и 38.

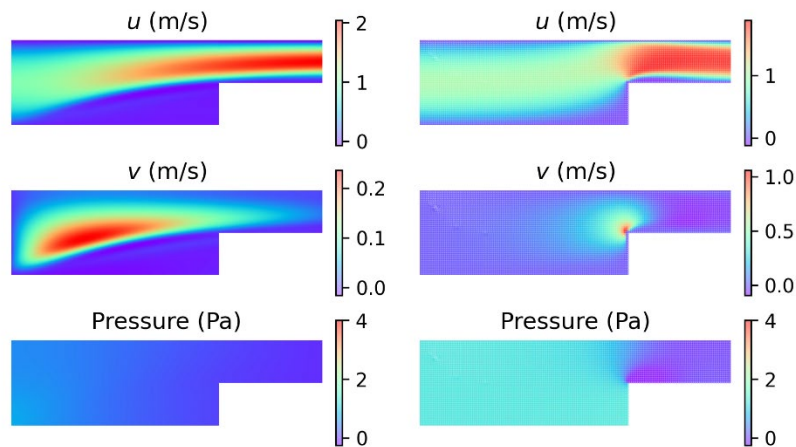


Рисунок 37. Расчет на 80 000 итераций слева - PINN, справа - Fluent для $Re \approx 500$.

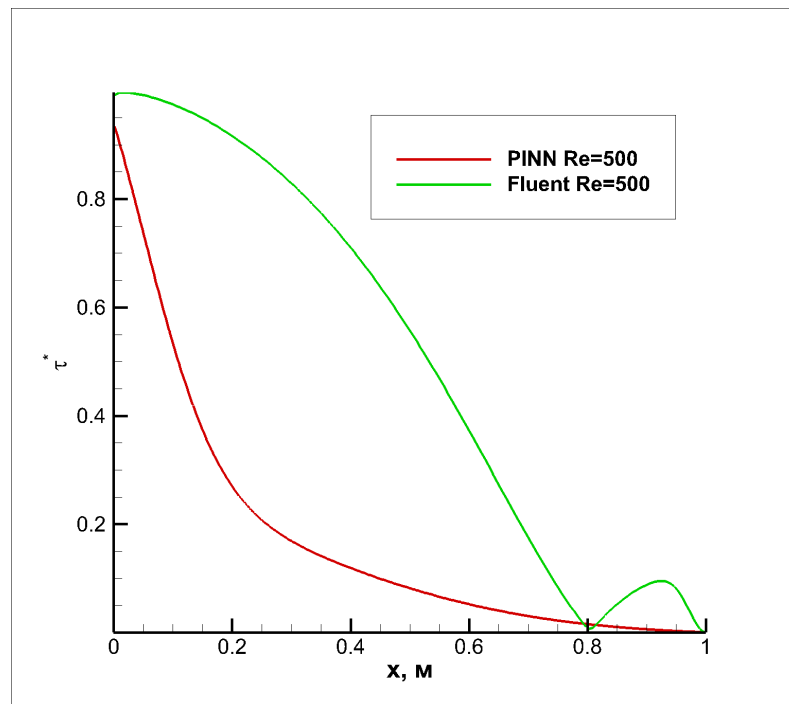


Рисунок 38. График безразмерного напряжения трения на нижней стенке при $Re = 500$ и 80 000 итераций.

Таблица 4.

Re	$u_{\%}, \%$	$v_{\%}, \%$	$p_{\%}, \%$	$\tau_{\%}^*, \%$
500	35	8	61	65

В ходе исследования были предприняты попытки коррекции данного поведения с использованием различных методов, включая введение регуляризации, нормализации, а также метода dropout в нейронную сеть, изменение количества слоев и нейронов, изменение размерности сетки и шага

обучения (learning rate). Однако, эти действия не привели к значимому улучшению результата. Отмечена лишь зависимость получаемых результатов от выбранной сетки.

В планах на будущее заложено проведение дополнительных исследований для нахождения решения этой проблемы. Вероятно, одной из причин такого поведения может быть особенность подхода к прогнозу решения уравнений Навье-Стокса с использованием тензора напряжений.

5. ОБРАТНЫЙ ПРОГНОЗ PINN

Физико-информированные нейронные сети не ограничиваются лишь прямым прогнозированием на основе предварительно заданных уравнений. Они также способны обеспечивать обратный прогноз (Inverse PINN), когда на основе имеющихся данных можно получить информацию о соответствующих уравнениях. Это, в частности, позволяет узнать значения коэффициентов или других неизвестных параметров уравнений. Несмотря на то, что наша нейронная сеть разрабатывается по аналогичному принципу, применяемый в данном случае подход к определению коэффициентов уравнения отличается тем, что он не основан на физических предположениях. Вместо этого мы используем метод обратного PINN, который по своей сути эквивалентен регрессионному анализу, то есть подбору коэффициентов некоторой зависимости, аппроксимирующей имеющиеся данные.

5.1 Постановка задачи

В качестве основы для обучения нейронной сети были использованы данные из [22]. Данное исследование направлено на изучение влияния геометрических и прочих параметров на течение и эпитаксию в прототипе нового реактора с помощью численного моделирования в пакете OpenFOAM. Данный тип исследования подразумевает выполнение большого объема ресурсоемких вычислений, и в случае моделирования всех процессов в полном объеме может привести к нецелесообразно большим временным затратам на получение результата. Таким образом, применение нейронных сетей для прогнозирования величин, представляющих интерес для исследователей, может существенно способствовать ускорению и упрощению вычислительных процессов.

CVD (chemical vapor deposition) реактор — это технологическое оборудование, которое используется для синтеза тонких пленок на поверхности различных материалов. Принцип работы CVD реактора состоит в том, что он использует газовые реагенты, которые разлагаются при высокой температуре, чтобы синтезировать кристалл на поверхности образца. Такие пленки могут быть из различных материалов, например, металлов, соединений металлов и неметаллов, полупроводников и т.д. Они используются в различных областях, таких как электроника, химическое оборудование, автомобилестроение и т.д.

Расчетная область, в которой проходят химические реакции, исследуемого реактора [22] показана на рисунке 39:

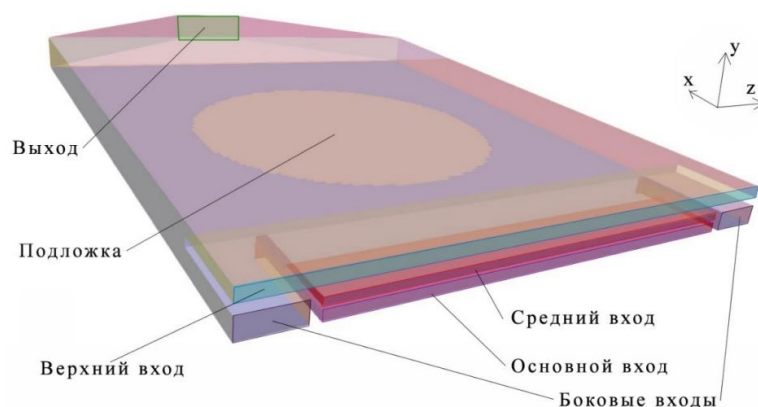


Рисунок 39. Расчетная область [22].

Для реализации нейронной сети был использован набор данных из работы [22], собранный в ходе приближенно 200 численных расчетов (CFD) для данного типа реактора. Исходный набор включал 28 параметров, однако после анализа корреляций было решено сократить их количество до 15, включающих в себя параметры геометрии реактора (7), скоростные показатели на входе (3), химический состав реагирующего вещества (4), а также один ключевой параметр, представляющий особый интерес для данного прогнозирования — среднееквадратичное значение скорости роста кристаллов вдоль линии диаметра длиной 200 мм, расположенной поперек потока, обозначенной как линия АВ на Рисунке 40.

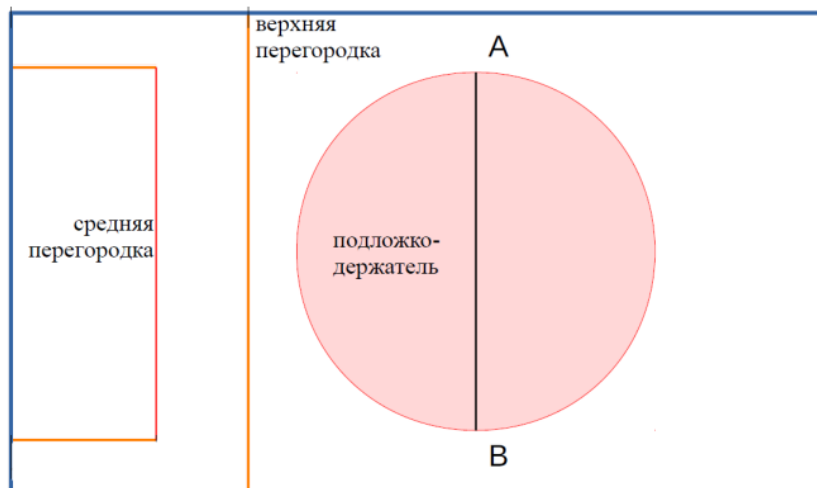


Рисунок 40. Схема реактора вид сверху [22].

В работе [22] исследуется CVD-реактор горизонтального типа, характеризующийся зональной подачей газов вдоль подложки. В частности, реактор имеет шесть зон подачи, что формирует структуру скорости роста вдоль подложки с пятью максимумами. Предполагается, что различие в числе зон подачи и максимумов может быть обусловлено особенностями гидродинамики потока. Иллюстрация такой структуры представлена на Рисунке 41, где границы подложки обозначены красной окружностью.

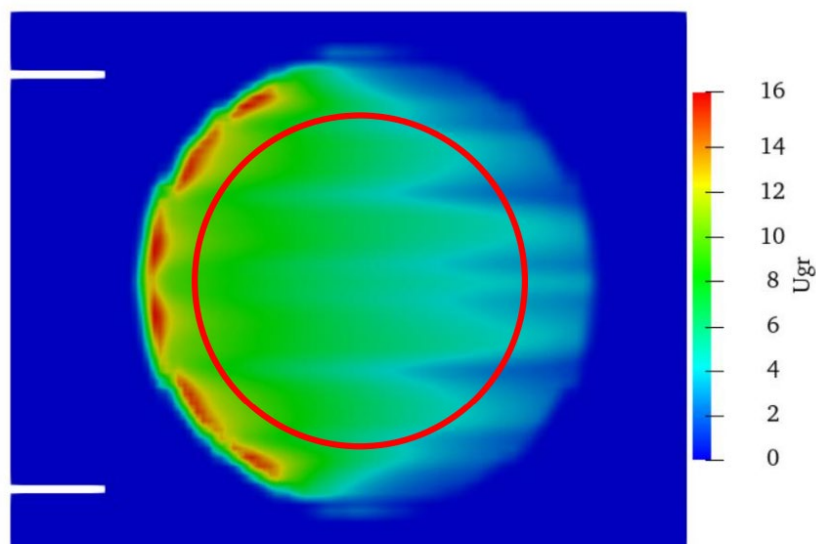


Рисунок 41. [22] Поле скорости роста кристалла на нижней стенке, $\mu\text{м}/\text{час}$

На Рисунке 42 представлено распределение скорости роста кристаллов по диаметру поперечного сечения подложки. Здесь можно более явно наблюдать наличие пяти максимумов.

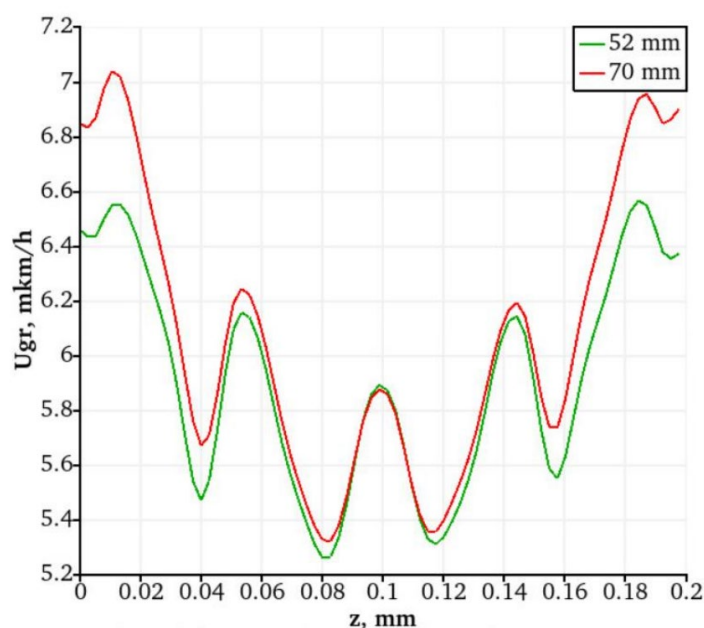


Рисунок 42. График скорости роста вдоль диаметра подложки поперек потока [22].

Было выдвинуто предположение, что данный график может быть описан суммой Гауссовых функций следующего вида:

$$y_1(z) = k_1 \frac{1}{\sqrt{\pi}} e^{-s_1(z-z_1)^2}$$

$$y(z) = \sum_{i=1}^5 y_i$$

, где k_i и s_i – это подлежащие определению нейронной сетью коэффициенты, x_i известные величины смещения. Предполагается, что распределения для всех расчетов имеют аналогичный вид и смещения сохраняют постоянство.

Принимая во внимание предположение о симметричности распределения скорости относительно нуля, можно прогнозировать не пять, а три коэффициента k_i и s_i устанавливая значения смещений $z_1=0$, $z_{2,3} = \pm 0.0445$, $z_{4,5} = \pm 0.088$.

На Рисунке 43 представлен график такого распределения, описываемого суммой Гауссовых функций при различных предустановленных коэффициентах.

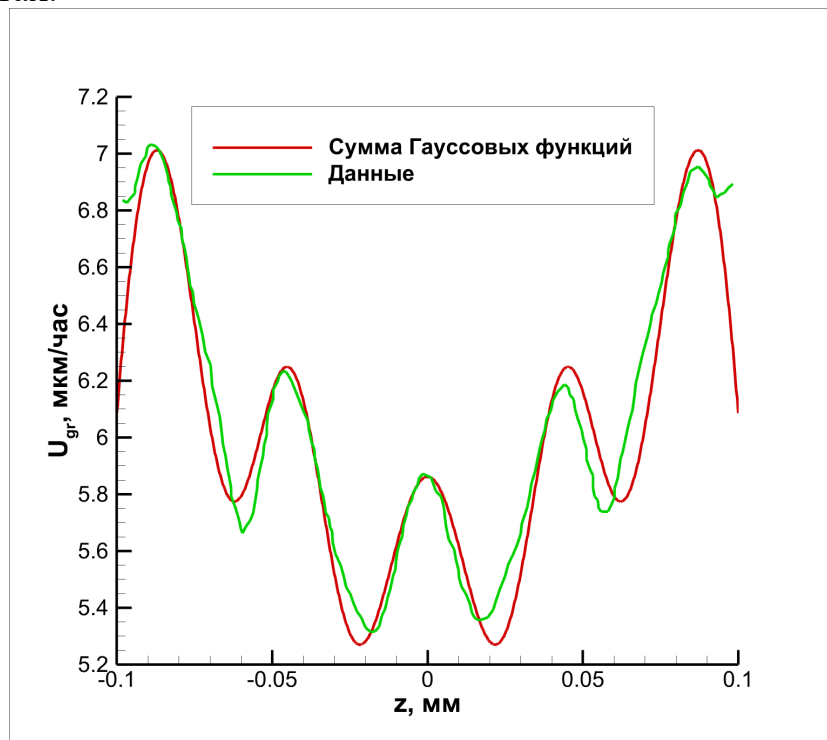


Рисунок 43. График суммы пяти Гауссовых функций при различных коэффициентах.

5.2 Структура нейронной сети

С использованием библиотеки PyTorch, на основе имеющихся 14 входных параметров была создана обратная физико-информированная нейронная сеть, коэффициенты в функции ошибок.

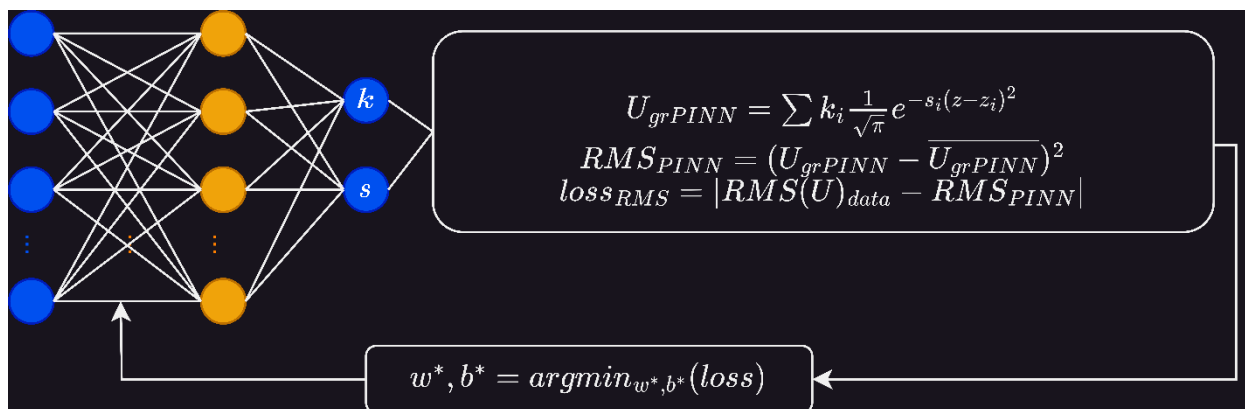


Рисунок 44. Архитектура нейронной сети PINN с обратным прогнозом.

Согласно представленной архитектуре нейросеть принимает на вход данные, прогнозирует тензоры коэффициентов размера (3, 1) и на их основе производит расчет среднеквадратичного значения относительно среднего вдоль z по формуле:

$$RMS(y(z)) = \sqrt{\frac{1}{0.2} \int_{-0.1}^{0.1} (y(z) - \overline{y(z)})^2 dz}$$

После чего происходит расчет ошибки спрогнозированного значения в сравнении с данными и корректировка весов и смещений нейронов с целью минимизации ошибки.

5.3 Результат работы программы

На предоставленных данных была обучена PINN нейронная сеть. При проверке использовалась валидационная выборка – случайно выбранные 10% от изначального набора данных, которые не использовались в процессе обучения. Результат прогноза в сравнении с валидационной выборкой, можно наблюдать в таблице:

Таблица 5.

$RMS(U_{gr})_{data}, \frac{\text{МКМ}}{\text{час}}$	$RMS(U_{gr})_{PINN}, \frac{\text{МКМ}}{\text{час}}$	Различие, %
0.2254	0.3963	75.846
0.415	0.3551	14.433
0.3906	0.3711	4.991
0.3747	0.3671	2.013
0.2962	0.3792	28.011
0.4495	0.3753	16.522
0.3309	0.3712	12.176
0.376	0.368	2.117
0.1748	0.3752	114.581
0.3789	0.3776	0.351

0.2843	0.3655	28.559
1.3528	0.3685	72.758
0.3241	0.3687	13.782
0.2403	0.3742	55.733
0.4196	0.3907	6.892
0.2557	0.3719	45.425

, где U_{gr} – скорость роста кристалла.

Архитектура нейросети PINN, применяемая в данном исследовании, продемонстрировала свою способность к прогнозированию коэффициентов на основе входных данных. Однако стоит отметить, что при валидации получилась значительная средняя ошибка, составляющая 35%. Во многом это можно объяснить ограниченностью доступного набора данных и недостаточным разбросом величин в нем. Также возможна дальнейшая оптимизация процесса обучения или корректировки структуры нейросети.

ЗАКЛЮЧЕНИЕ

В рамках представленной работы была исследована возможность применения нейросетевых технологий в области вычислительной гидродинамики. На основе проведенного анализа было сформулировано обоснованное предположение о наиболее перспективном направлении развития в данной области, которым, как было выявлено, являются физико-информированные нейронные сети.

Основываясь на обзоре существующих работ и направлений развития в данной области, были разработаны две программы, обеспечивающие возможность проведения прогнозирования на основе применения нейронных сетей в контексте ламинарных задач вычислительной гидродинамики.

Первый программа, реализует физико-информированную нейронную сеть, проводя прямое прогнозирование ламинарного течения. В рамках данной модели физические уравнения Навье-Стокса и установленные граничные условия играют центральную роль в процессе обучения нейронной сети, также предусмотрена возможность ввода дополнительных данных.

Программа была применена для расчета течения в канале с внезапным сужением. Были проанализированы и исследованы параметры, которые оказывают наибольшее влияние на процесс обучения физико-информированной нейронной сети в рамках решаемой задачи.

Результаты, полученные с помощью физико-информированной нейросети с прямым прогнозом для чисел Рейнольдса, не превышающих 100, можно считать удовлетворительными. Однако при числах Рейнольдса, превышающих 100, результаты отдаляются от физических ограничений и не могут быть приняты в качестве достоверных расчетов. Поиск путей решения этой проблемы представляет собой важное направление для будущих исследований.

Вторая программа представляет собой инструмент для проведения обратного прогнозирования. На основании имеющихся 200 CFD-расчетов в CVD-реакторе горизонтального типа, была разработана нейросеть, осуществляющая подбор коэффициентов пяти Гауссовых функций, способных имитировать распределение скорости роста вдоль поперечного сечения подложки, на которой происходит рост кристалла. Точность работы данной сети оказалась относительно невысокой, со средней ошибкой валидации около 35%, что может быть объяснено ограниченностью и недостаточной разнообразностью представленных данных.

Проведенная работа демонстрирует важность и актуальность развития физико-информированных нейронных сетей в области вычислительной гидродинамики. В ходе этого исследования были определены возможности и ограничения данного подхода, а также выявлены потенциальные направления для дальнейших исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Rabault J., Kuchta M., Jensen A., Réglade U., Cerardi N. Machine Learning for Fluid Mechanics, 2019.
2. Ricardo Vinuesa, Steven L. Brunton. Enhancing Computational Fluid Dynamics with Machine Learning, 2021.
3. Chengping Rao, Hao Sun, Yang Liu. Physics-informed deep learning for incompressible laminar flows, 2020.
4. George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, Liu Yang. Physics-informed machine learning, 2021.
5. Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi and Francesco Piccialli. Scientific Machine Learning through Physics-Informed Neural Networks: Where we are and What's next, 2022.
6. Natasha Dropka, Martin Holena. Application of Artificial Neural Networks in Crystal Growth of Electronic and Opto-Electronic Materials, 2020.
7. Kentaro Kutsukake, Yuta Nagai, Tomoyuki Horikawa, Hironori Banba. Real-time prediction of interstitial oxygen concentration in Czochralski silicon using machine learning, 2020.
8. Kentaro Kutsukake, Yuta Nagai, Hironori Banba. Virtual experiments of Czochralski growth of silicon using machine learning: Influence of processing parameters on interstitial oxygen concentration, 2022.
9. Yosuke Tsunooka, Nobuhiko Kokubo, Goki Hatasa, Shunta Harada, Miho Tagawa, Toru Ujiharaab. High-speed prediction of computational fluid dynamics simulation in crystal growth, 2018.
10. Xiaofang Qi, Wencheng Ma, Yifan Dang, Wenjia Su, Lijun Liu. Optimization of the melt/crystal interface shape and oxygen concentration during the Czochralski silicon crystal growth process using an artificial neural network and a genetic algorithm, 2020.
11. Yifan Dang, Lijun Liu, Zaoyang Li. Optimization of the controlling recipe in quasi-single crystalline silicon growth using artificial neural network and genetic algorithm, 2019.
12. Donglin Chen, Xiang Gao, Chuanfu Xu, Shizhao Chen, Jianbin Fang, Zhenghua Wang, Zheng Wang. FlowGAN: A Conditional Generative Adversarial Network for Flow Prediction in Various Conditions, 2020.
13. ShuranYe, ZhenZhang, Xudong Song, YiweiWang, YaosongChen, Chenguang Huang. A flow feature detection method for modeling pressure distribution around a cylinder in non-uniform flows by using a convolutional neural network, 2020.
14. Masaki Morimoto, Kai Fukami Kai Zhang, Aditya G. Nair, Koji Fukagata. Convolutional neural networks for fluid flow analysis: toward effective metamodeling and low-dimensionalization, 2021.

15. Natasha Dropka, Martin Holena, Stefan Ecklebe, Christiane Frank-Rotsch, Jan Winkler. Fast forecasting of VGF crystal growth process by dynamic neural networks, 2019.
16. Jing Zhang, Qinwei Tang, Ding Liu. Research Into the LSTM Neural Network-Based Crystal Growth Process Model Identification, 2019.
17. Shengze Cai, Zhiping Mao, Zhicheng Wang, Minglang Yin, George Em Karniadakis. Physics-informed neural networks (PINNs) for fluid mechanics: A review, 2021.
18. Shengze Cai, Zhicheng Wang, Frederik Fuest, Young-Jin Jeon, Callum Gray, George Em Karniadakis. Flow over an espresso cup: Inferring 3D velocity and pressure fields from tomographic background oriented schlieren videos via physics-informed neural networks, 2021.
19. Shengze Cai, Zhicheng Wang, Frederik Fuest, Young-Jin Jeon, Callum Gray, George Em Karniadakis. Supplementary materials: Flow over an espresso cup: Inferring 3D velocity and pressure fields from tomographic background oriented schlieren videos via physics-informed neural networks, 2021.
20. Stefano Markidis. «The Old and the New: Can Physics-Informed Deep-Learning Replace Traditional Linear Solvers? », 2021.
21. Xu Liu, Xiaoya Zhang, Wei Peng, Weien Zhou, Wen Yao. A novel meta-learning initialization method for physics-informed neural networks, 2021.
22. Е.А. Калаушина. Численное моделирование газофазной эпитаксии из металлоорганических соединений в пакете OpenFOAM, магистерская диссертация, 2021, 65 стр.