

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра информатики и программирования

**ПРОГНОЗ ВЫСОТЫ СНЕЖНОГО ПОКРОВА ПО ДАННЫМ  
ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ NASA В ПРЕДЕЛАХ  
РАЙОНОВ ДВУХ МЕТЕОСТАНЦИЙ**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студентки 4 курса 441 группы

направления 02.03.03 Математическое обеспечение и администрирование  
информационных систем

факультета компьютерных наук и информационных технологий

Заяц Наталии Алексеевны

Научный руководитель

ст. преподаватель

\_\_\_\_\_

Лапшева Е.Е.

Зав. кафедрой

к.ф.-м.н., доцент

\_\_\_\_\_

Огнева М. В.

Саратов 2026

## ВВЕДЕНИЕ

**Актуальность темы.** В условиях наблюдаемых климатических изменений режим снежного покрова существенно трансформируется, что усиливает неопределённость при гидрологическом прогнозировании [1]. Традиционная сеть метеостанций обеспечивает лишь точечные измерения высоты снежного покрова (ВСП) и не отражает его пространственной неоднородности, обусловленной рельефом, растительностью и ветровым режимом [2]. Интеграция спутниковых данных NASA MODIS с наземными измерениями посредством методов машинного обучения позволяет преодолеть это ограничение и получить количественные пространственно распределённые оценки ВСП [3, 4]. Особую значимость данная задача приобретает для равнинных районов с засушливым климатом юго-востока Европейской России – степной и полупустынной зон Саратовской области, где неустойчивость снежного покрова и его высокая межгодовая изменчивость создают повышенную нагрузку на системы гидрологического прогнозирования.

**Цель бакалаврской работы** – разработать модель прогнозирования высоты снежного покрова на основе данных дистанционного зондирования Земли (NASA) и наземных измерений метеостанций.

Поставленная цель определила **следующие задачи**:

1. Провести обзор научных источников по теме исследования.
2. Изучить существующие методы прогнозирования высоты снежного покрова – традиционные и основанные на машинном обучении.
3. Дать физико-географическую характеристику районов исследования – метеостанций Балашов и Александров Гай.
4. Собрать данные дистанционного зондирования NASA (MODIS MOD10A1), реанализа ERA5 и наземных метеонаблюдений за период 2009-2024 гг.
5. Выполнить предобработку данных: конвертацию, пространственную привязку, извлечение значений по точкам метеостанций и объединение

в единый датасет.

6. Провести разведочный анализ данных: оценить качество, полноту, распределения признаков и корреляции; отобрать информативные предикторы.
7. Выбрать и обучить модели машинного обучения для прогнозирования высоты снежного покрова.
8. Оценить качество моделей машинного обучения по метрикам (MAE, ) и выбрать наилучшую.
9. Выполнить оптимизацию гиперпараметров наилучшей модели машинного обучения с целью повышения точности прогноза.
10. Выбрать и обучить модели нейронных сетей для решения той же задачи.
11. Оценить качество нейросетевых моделей и выбрать наилучшую.
12. Выполнить оптимизацию наилучшей нейросетевой модели.
13. Провести сравнительный анализ лучших моделей машинного обучения и нейронных сетей; сформулировать выводы о применимости каждого подхода.

**Методологические основы** прогнозирования высоты снежного покрова по данным дистанционного зондирования Земли и методов машинного обучения представлены в работах П.П. Кузьмина, Д.К. Холл и Дж.А. Риггса, Е.В. Чурюлина и др., С. Гараи-Манеша и др., Г.В. Айзеля, А. Резы и др., Д.Х. Канга и др.

**Теоретическая значимость бакалаврской работы.** Работа систематизирует и сравнивает традиционные статистические, ансамблевые и нейросетевые методы прогнозирования высоты снежного покрова применительно к равнинным аридным и полуаридным территориям, для которых подобные исследования единичны. Полученные результаты расширяют представления о применимости различных классов моделей машинного обучения в условиях выраженной прерывистости снежного покрова и высокой доли нулевых значений целевой переменной.

**Практическая значимость бакалаврской работы.** Разработанная модель и методика совместной обработки спутниковых данных NASA MODIS, реанализа ERA5 и наземных наблюдений могут быть использованы в системах гидрологического и агрометеорологического прогнозирования Саратовской области, в частности для оценки высоты снежного покрова на территориях, не охваченных плотной сетью наземных станций. Разработанные программные модули для автоматизированного скачивания, конвертации и пространственной привязки данных дистанционного зондирования могут применяться при решении аналогичных задач мониторинга снежного покрова в других регионах.

**Структура и объём работы.** Бакалаврская работа состоит из введения, 3 разделов, заключения, списка использованных источников и 11 приложений. Общий объём работы – 103 страницы, из них 48 страниц – основное содержание, включая 2 рисунков и 3 таблиц, список использованных источников информации – 30 наименований.

## **КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Первый раздел «Теоретические основы прогнозирования высоты снежного покрова»** посвящен обзору теоретических и методологических основ исследования.

В разделе рассмотрены основные направления исследования снежного покрова – от классических работ российских гидрологов и климатологов до современных методов интеграции спутниковых данных NASA MODIS с алгоритмами машинного обучения. Дана физико-географическая характеристика районов двух метеостанций – Балашов (лесостепная зона) и Александров Гай (полупустынная зона) – и показаны принципиальные различия в режиме снежного покрова между ними. Задача прогнозирования высоты снежного покрова формализована как задача регрессии; рассмотрены основные группы алгоритмов машинного обучения (линейные модели, ансамбли деревьев решений, нейронные сети), метрики оценки качества

прогноза (MAE, MASE, ) и причины, по которым относительные метрики (MAPE, SMAPE, WAPE) непригодны при значительной доле нулевых наблюдений, а также методы борьбы с переобучением. Сделан вывод о целесообразности применения ансамблевых методов машинного обучения и нейросетевых моделей для решения поставленной задачи в условиях равнинной аридной территории.

**Второй раздел «Обработка и разведочный анализ данных»** посвящен самостоятельной разработке программных средств для сбора и обработки данных дистанционного зондирования, их предобработке и разведочному анализу.

Для решения поставленных задач созданы специализированные скрипты на языке Python. Реализован полный цикл обработки: авторизованное пакетное скачивание продукта MOD10A1 (NDSI) с портала NASA NSIDC за период 2009-2024 гг., конвертация растровых файлов из формата HDF4 в GeoTIFF, пространственное извлечение значений снежного индекса по точкам метеостанций Балашов и Александров Гай с использованием буферных зон радиусом 1000 м. Дополнительно обработаны данные реанализа ERA5 (суммарный, поверхностный и подземный сток) в формате NetCDF, а также приведены к единому CSV-формату разнородные текстовые файлы наземных наблюдений (температура, осадки, высота снега, снегосъёмки) с автоматическим определением кодировки. В результате объединения всех источников сформирован датасет объёмом 239 360 записей и 100 признаков за период 1936-2025 гг.

На этапе разведочного анализа выполнена агрегация метеорологических данных к суточной периодичности (осреднение температур, влажности, давления, ветра; суммирование осадков) для согласования с периодичностью спутниковых и снегомерных измерений. Проведена очистка данных: удалены пустые и избыточные столбцы, дублирующие идентификаторы и временные метки, а также признаки с долей пропусков >95 %. Обработка пропусков выполнена с помощью специальных

числовых маркеров (-1 для высоты облачности, -999 для температуры почвы, -1.5 для NDSI), что позволило алгоритмам машинного обучения интерпретировать отсутствие данных как отдельное состояние. Для устранения мультиколлинеарности из каждой пары признаков с коэффициентом корреляции удалён один, что сократило число предикторов со 100 до 19. Визуализация распределений выявила правостороннюю асимметрию осадков и высоты снега, бимодальность NDSI, а также естественную близость к норме температурных и влажностных рядов. Применение метода главных компонент показало, что 13 компонент объясняют 95 % дисперсии исходных признаков. Дополнительно созданы новые признаки: номер месяца, день года и лаговое значение высоты снежного покрова (сдвиг на одни сутки). По итогам анализа сформированы три версии итогового датасета – с исходными признаками, с PCA-компонентами и с лаговым признаком – что позволит в дальнейшем оценить влияние различных представлений данных на точность прогнозных моделей. Все этапы обработки и анализа подтвердили готовность данных к построению прогностических алгоритмов.

**Третий раздел «Обучение моделей машинного обучения прогнозированию высоты снежного покрова»** посвящен построению, обучению и сравнительной оценке моделей машинного обучения и нейронных сетей для прогнозирования высоты снежного покрова.

Прежде чем перейти к систематическому сравнению алгоритмов, была проведена серия экспериментов по выбору оптимального набора входных признаков. Рассмотрены три варианта: исходные метеорологические и спутниковые признаки без преобразований, признаки после снижения размерности методом главных компонент (PCA) и признаки с добавлением лагового значения высоты снежного покрова за предыдущий день.

В результате оптимальным был признан подход на основе исходных метеорологических и спутниковых признаков без преобразований, обеспечивающий наилучшее сочетание точности и практической

применимости для месячного горизонта прогноза.

Систематическое сравнение алгоритмов регрессии выстраивалось от простых моделей к сложным. Линейная регрессия показала наиболее низкое качество ( $R^2 = 0,32$ , MAE = 2,80 см) ввиду неспособности описывать пороговые и мультипликативные взаимодействия температуры и осадков, определяющие накопление снега. Полиномиальная регрессия второй степени улучшила результат ( $R^2 = 0,46$ , MAE = 1,80 см) за счёт учёта парных взаимодействий, однако повышение степени до 3-4 приводило к катастрофическому переобучению. Ансамбли деревьев решений продемонстрировали значительно более высокое качество: одиночное дерево решений полностью переобучалось (Train = 1,000 при Test = 0,811), случайный лес после подбора гиперпараметров (GridSearchCV, TimeSeriesSplit) устранил эту проблему, достигнув Test = 0,879 (MAE = 2,41 см). Наилучший результат среди одиночных алгоритмов показал XGBoost после оптимизации learning\_rate, max\_depth, n\_estimators и регуляризации: Test = 0,887, MAE = 2,37 см. LightGBM (Test = 0,878) и CatBoost (Test = 0,843) показали близкие, но несколько более низкие результаты. SVR с оптимальными параметрами оказался наименее эффективным среди всех алгоритмов ( $R^2 = 0,716$ , MAE = 4,00 см) из-за выраженной правосторонней асимметрии распределения ВСП. Опробованный двухэтапный подход «классификатор + регрессор» также уступил одноэтапной регрессии (MAE = 4,05 см) вследствие систематического смещения обучающей выборки регрессора.

Наилучшее качество среди всех рассмотренных моделей обеспечил стекинг-ансамбль, объединивший прогнозы случайного леса, XGBoost и LightGBM с помощью мета-модели – гребневой регрессии (Ridge), обученной на основе 7-кратной кросс-валидации, исключающей утечку информации из обучающей выборки в мета-уровень: Test = 0,898, MAE = 2,20 см. Выбор именно этих трёх базовых алгоритмов обусловлен их взаимодополняемостью: случайный лес устойчив вблизи нулевого порога, XGBoost точнее при умеренном устойчивом снежном покрове, а LightGBM

эффективнее в переходные периоды – совместно они перекрывают основные режимы состояния снежного покрова в исследуемом регионе.

Параллельно с классическими алгоритмами машинного обучения проведён эксперимент с пятью нейросетевыми архитектурами – LSTM, GRU, Bi-LSTM, CNN-1D и CNN-LSTM. Их качество оказалось значительно ниже ансамблевых методов: Test составил от -0,120 (GRU) до 0,182 (CNN-LSTM – лучшая из рассмотренных архитектур), MAE – от 3,08 до 3,60 см. Такое отставание объясняется недостаточным для глубоких сетей объёмом обучающей выборки, высокой долей нулевых значений целевой переменной, при которой нейросети «застревают» в тривиальном решении, а также прерывистым характером снежного покрова на исследуемой территории, разрушающим связность временного ряда.

Итоговое сравнение всех рассмотренных моделей на тестовой выборке, не участвовавшей в обучении и настройке гиперпараметров, подтвердило, что ансамблевые методы машинного обучения, и в первую очередь стекингый ансамбль, обеспечивают наилучшую точность и устойчивость прогноза высоты снежного покрова для равнинных аридных территорий, превосходя как линейные модели, так и нейросетевые архитектуры при ограниченном объёме наземных наблюдений. Достигнутый уровень точности (MAE = 2,20 см,  $R^2 = 0,898$ ) сопоставим с результатами аналогичных исследований для других полуаридных регионов и близок к физически обоснованному нижнему пределу точности месячного прогноза, определяемому непредсказуемостью атмосферных процессов на данном горизонте.

## **ЗАКЛЮЧЕНИЕ**

В данной выпускной квалификационной работе была разработана модель прогнозирования высоты снежного покрова на основе данных дистанционного зондирования NASA и наземных измерений метеостанций Балашов и Александров Гай за период 2009-2024 гг. Все поставленные задачи выполнены, цель достигнута.

В ходе работы из 100 исходных переменных отобраны 19 наиболее информативных признаков: температурные показатели, осадки, индекс NDSI и данные реанализа ERA5. Реализованы и обучены модели машинного обучения и нейросетевые архитектуры с последующей оптимизацией гиперпараметров лучших из них. Сравнительный анализ показал, что ансамблевые методы – в частности, градиентный бустинг и случайный лес – обеспечивают наилучшую точность прогноза и устойчивость в различных природных условиях, превосходя нейросетевые модели при ограниченном объеме обучающей выборки. Совместное использование спутниковых данных NASA и наземных наблюдений позволило существенно повысить качество прогноза по сравнению с применением каждого источника в отдельности.

Полученные результаты подтверждают применимость методов машинного обучения для прогнозирования высоты снежного покрова на равнинных аридных территориях и могут быть использованы в системах гидрологического прогнозирования Саратовской области.

### **Основные источники информации:**

1. Кузьмин П.П. Физические свойства снежного покрова. – Л.: Гидрометеиздат, 1957. – 180 с.
2. Hall D.K., Riggs G.A., Salomonson V.V. MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 61 [Электронный ресурс] / NASA NSIDC DAAC, 2016. – DOI: 10.5067/MODIS/MOD10A1.061.
3. Reza A. [et al.] Leveraging advanced deep learning and machine learning approaches for snow depth prediction // Remote Sensing Applications: Society and

Environment. – 2024. – DOI: 10.1016/j.rsase.2024.101427.

4. Kang D.H. [et al.] Comparison of machine-learning-based snow depth estimates // Remote Sensing. – 2022. – Vol. 14, № 12. – P. 2800. – DOI: 10.3390/rs14122800.

5. Чурюлин Е.В. [и др.] Анализ характеристик снежного покрова по спутниковым и модельным данным // Гидрометеорологические исследования и прогнозы. – 2018. – № 2 (368). – С. 120–143.

6. Gharaei-Manesh S., Fathzadeh A., Taghizadeh-Mehrjardi R. Comparison of artificial neural network and decision tree models in estimating spatial distribution of snow depth in a semi-arid region of Iran // Cold Regions Science and Technology. – 2016. – Vol. 122. – P. 26–35. – DOI: 10.1016/j.coldregions.2015.11.004.

7. Айзель Г.В. Применение методов машинного обучения для моделирования толщины снежного покрова // Лёд и Снег. – 2017. – Т. 57, № 1. – С. 34–44. – DOI: 10.15356/2076-6734-2017-1-34-44.