

**НАЦИОНАЛНА СПОРТНА АКАДЕМИЯ „ВАСИЛ ЛЕВСКИ“
КАТЕДРА „ТЕОРИЯ НА СПОРТА“**

АВТОРЕФЕРАТ

на дисертационен труд за присъждане на научната степен
„ДОКТОР НА НАУКИТЕ“

в професионално направление 7.6. „Спорт“

**ИНТЕГРАЛНА МЕТОДИКА ЗА ОПТИМИЗИРАНЕ НА
ТРЕНИРОВЪЧНОТО НАТОВАРВАНЕ ЧРЕЗ
ГЕНЕРАТИВНИ МОДЕЛИ НА
ИЗКУСТВЕН ИНТЕЛЕКТ**

Проф. Михаил Ивайлов Кончев, доктор

София, 2026

ДАННИ ЗА ДИСЕРТАЦИОННИЯ ТРУД

Дисертационният труд е разработен в обем 281 страници. Съдържа 32 фигури и 16 таблици, както и 4 приложения.

Публичната защита на дисертационният труд ще се състои на 17.03.2026 г. в зала А-3 на НСА „Васил Левски“.

РЕЦЕНЗЕНТИ

проф. Свилен Емилов Нейков, ДН

проф. Даниела Станимирова Дашева, ДН

проф. Димитър Георгиев Велев, доктор

СПИСЪК НА ИЗПОЛЗВАНИТЕ СЪКРАЩЕНИЯ И ОЗНАЧЕНИЯ

- AI – изкуствен интелект (Artificial Intelligence).
- Generative AI – генеративни модели на изкуствения интелект за синтез на нови данни/структури.
- GAN – генеративна състезателна мрежа (Generative Adversarial Network): генератор и дискриминатор в минимакс постановка.
- VAE – вариационен автоенкодер (Variational Autoencoder): вероятностна латентна репрезентация и KL-регуляризация.
- KL-дивергенция – информационна мярка за разлика между две вероятностни разпределения (Kullback–Leibler).
- САГ – средна абсолютна грешка: средно отклонение между реални и генерирани стойности по режими/зони на натоварване.
- $\text{VO}_{2\max}$ – максимална кислородна консумация ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) като показател за аеробен капацитет.
- Zone 1–Zone 5 – зони на натоварване (време в часове/минути), описващи разпределението на тренировъчния обем по енергийни режими.
- Control_days / Main_days – оставащи дни до контролно/основно състезание като контекстни (условни) променливи.

СЪДЪРЖАНИЕ

1. УВОД
2. АКТУАЛНОСТ И ЗНАЧИМОСТ НА ИЗСЛЕДВАНЕТО
3. МЕТОДИЧЕСКИ ОСНОВИ НА ИЗСЛЕДВАНЕТО
4. МЕТОДОЛОГИЧНА ОСНОВА И ИЗСЛЕДОВАТЕЛСКИ ДИЗАЙН
5. ТЕОРЕТИЧНА РАМКА: ТРЕНИРОВЪЧНОТО НАТОВАРВАНЕ КАТО ЗАДАЧА ЗА ОПТИМИЗИРАНЕ
6. ОПТИМИЗИРАНЕ НА ТРЕНИРОВЪЧНОТО НАТОВАРВАНЕ ЧРЕЗ ГЕНЕРАТИВНИ МОДЕЛИ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ
7. ИНТЕГРАЛНА МЕТОДИКА ЗА ОПТИМИЗИРАНЕ НА ТРЕНИРОВЪЧНОТО НАТОВАРВАНЕ ЧРЕЗ ГЕНЕРАТИВНИ МОДЕЛИ НА ИЗКУСТВЕН ИНТЕЛЕКТ
8. ЕМПИРИЧНА БАЗА, ДАННИ И ПАРАМЕТРИЗАЦИЯ НА ГЕНЕРАТИВНИЯ МОДЕЛ
9. ПРОЕКТ НА ГЕНЕРАТИВНИЯ МОДЕЛ И ВЕРИФИКАЦИЯ
10. ЕКСПЕРИМЕНТАЛНО ПРИЛОЖЕНИЕ: МЛАДЕЖКИ НАЦИОНАЛЕН ОТБОР ПО БИАТЛОН (СЕЗОН 2023-2024)
11. РЕЗУЛТАТИ И АНАЛИЗ
12. ЗАКЛЮЧЕНИЕ
13. НАУЧНИ И ПРИЛОЖНИ ПРИНОСИ

1. УВОД

Оптимизирането на тренировъчните натоварвания е сред най-сложните методически задачи в теорията и методиката на спортната тренировка. Тренировъчният процес се характеризира с висока степен на динамичност, многокомпонентност и наличие на факторни и областни ограничения, произтичащи от функционалните възможности на спортиста, етапа от подготовката и логиката на периодизацията.

В практиката натоварването се разглежда като комплекс от параметри, които следва да бъдат количествено определени и управлявани така, че да се постигне максимален функционален ефект при запазване на физиологичната стабилност и контрол на риска от преумора и травматизъм. Това предполага формулиране на критерий за оптималност (целева функция), избор на управляващи параметри и прилагане на алгоритми за търсене на оптимум в допустимото пространство на решенията.

Развитието на генеративните модели на изкуствения интелект създава възможност тренировъчните програми да бъдат разглеждани като реализации на многоизмерно вероятно разпределение, извлечено от експертно разработени и емпирично валидирани планове. По този начин се открива подход за глобално изследване и синтез на тренировъчни натоварвания, последвано от контрол и статистическа верификация.

2. АКТУАЛНОСТ И ЗНАЧИМОСТ НА ИЗСЛЕДВАНЕТО

Актуалността на дисертационния труд произтича от обективната необходимост тренировъчният процес да бъде управляван като динамична и многокритериална система, в която решенията за натоварването следва да отчитат ограниченията, адаптацията и периодизацията на тренировъчния процес. В условията на високо спортно майсторство дори малки промени в обема и интензивността могат да имат съществено въздействие върху функционалните и структурните реакции на организма.

Съвременето ни е белязано от бурно развитие и масово внедряване на изкуствен интелект, включително генеративни модели, които се утвърждават като водещо направление в света и в България. В спортната наука това развитие създава предпоставки за преход от емпирично и частично оптимизиране към систематично, данни-ориентирано и възпроизводимо управление на тренировъчните натоварвания.

Значимостта на изследването се определя от предложението за интегрална методика, която обединява: (1) класическите постановки на спортната методика (величина, характер, насоченост; доза-ефект; натоварване и възстановяване), (2) математическо моделиране и оптимизационни алгоритми, и (3) генеративни модели на изкуствения интелект, подложени на статистическа и логическа верификация.

В дисертационния труд тренировъчната подготовка е разгледана в рамките на съвременните системи на спортна подготовка, където високите постижения се формират като резултат от сумарното влияние на генотипни, фенотипни, материално-технически, управленски и научно-технологични фактори. В този контекст тренировъчното натоварване се явява основният управляващ фактор, чрез който се реализира целенасочено въздействие върху функционалните адаптации и спортната форма.

Актуалността на тематиката се усилва от нарастващата достъпност на данни и от възможностите на Generative AI да синтезира структури, които съхраняват

закономерностите на емпиричните програми, но позволяват и систематично изследване на допустимото пространство на управляващите параметри. Това създава предпоставки за преминаване от локални и частични решения към възпроизводима интегрална методика, в която оптимизационните решения са подкрепени от статистическа оценка и логическа верификация.

За българската спортна практика значимостта на предложената в дисертационния труд интегрална методика се изразява в приложимостта ѝ при ограничени извадки и в контекста на младежкия спорт, където необходимостта от прецизно управление на натоварването е съчетана с по-висок риск от методически грешки. Верификационният модул с формализирани правила ограничава генерирането на физиологично неадекватни комбинации и поддържа съответствие с принципите на спортната методика.

3. МЕТОДИЧЕСКИ ОСНОВИ НА ИЗСЛЕДВАНЕТО

Целта на дисертационния труд е да се предложи интегрална методика за оптимизиране на тренировъчния процес чрез генеративни модели на изкуствения интелект, като предложената методика се валидира с данни от младежкия национален отбор по биатлон.

3.1. Научна теза

Дисертационният труд защитава тезата, че интегрирането на генеративни модели на изкуствения интелект в планирането и управлението на тренировъчния процес позволява систематична, адаптивна и многокритериална оптимизация на натоварванията в цикличните спортове.

3.2. Задачи на изследването

1. Да се направи анализ на теоретичните основи на тренировъчния процес като обект на оптимизиране и да се систематизират ключовите концепции за планиране, структуриране и управление на тренировъчните натоварвания.

2. Да се направи критичен анализ на съществуващите методи и алгоритми за оптимизиране, приложими към тренировъчните натоварвания, включително методи на системно-структурен анализ, математическо моделиране и адаптивни алгоритми за многокритериално оптимизиране.

3. Да се анализира приложимостта на генеративните модели на изкуствения интелект за оптимизиране на тренировъчните натоварвания и да се оценят техните предимства и ограничения.

4. Да се предложи концептуален модел за интегриране на генеративните модели в тренировъчния процес, включващ методологична рамка за събиране, структуриране и обработка на данни и етапи на адаптивно оптимизиране.

5. Да се разработи интегрален модел за оптимизиране на тренировъчното натоварване чрез генеративни модели на изкуствения интелект.

6. Да се приложи емпирично разработеният модел с данни от биатлона и да се генерира тренировъчна програма в реални условия.

7. Да се направи оценка на приложимостта на предложения интегрален модел чрез математико-статистически методи.

3.3. Обект и предмет на изследването

Обект на изследването са тренировъчните натоварвания в цикличните спортове като процес на планиране, адаптация и оптимизиране.

Предмет на изследването е разработването и валидирането на интегрална методика за оптимизация на тренировъчния процес чрез генеративни модели на изкуствения интелект, включително алгоритми за адаптивно управление на натоварванията, многокритериална оптимизация и верификация чрез математико-статистически методи.

4. МЕТОДОЛОГИЧНА ОСНОВА И ИЗСЛЕДОВАТЕЛСКИ ДИЗАЙН

Методологичната основа на изследването съчетава системно-структурен подход, математическо моделиране и оптимизация, както и генеративни модели на изкуствения интелект, предназначени за синтез и селекция на тренировъчни сценарии.

4.1. Методи на изследването

Приложените методи са организирани в три основни групи:

- Теоретико-методологичен анализ: систематичен преглед и критичен анализ на теории и алгоритми за оптимизиране на тренировъчните натоварвания и генеративни модели (VAE, GAN и др.).

- Експертна оценка: участие на експерти при генериране и структуриране на тренировъчните програми за обучаваща и тестова извадка, както и при дефиниране на входните параметри (VO_{2max_input} , $Control_days$, $Main_days$) и методически ограничения.

- Математико-статистически методи: вариационен анализ; квантилен анализ и стратификация по VO_{2max} (Q1-Q4); средна абсолютна грешка (САГ) и KL-дивергенция за оценка на структурното сходство; логически правила за контрол на вътрешната съгласуваност; тест Shapiro-Wilk за нормалност; тест Wilcoxon за зависими извадки; размер на ефекта (r) и анализ на зависимости чрез корелации и условни вероятности.

4.2. Дизайн на изследването

Експерименталната работа е реализирана чрез последователност от научноизследователски етапи: теоретичен етап (литературен обзор и методологична основа), разработване на интегрална методика, експертно генериране на тренировъчни програми за обучение на модела, статистическа оценка на резултатите и обобщение/дискусия.

Дизайнът на изследването е ориентиран към построяване на цикъл за управление, в който: (1) емпиричните данни от експертно планирани програми се структурират и стандартизират; (2) генеративният модел се обучава и валидира върху отделен набор от данни; (3) генерираните програми се подлагат на количествена оценка (САГ, KL-дивергенция) и логическа/физиологична проверка; (4) резултатите се използват за настройка на параметрите на модела и за формулиране на приложни препоръки.

Емпиричната база включва 200 индивидуални тренировъчни програми с различна продължителност, представени като 44 694 дневни наблюдения. Всеки дневен запис съдържа входни показатели (VO_{2max_input} и контекстни променливи до състезание),

както и изходни показатели – време по зони на натоварване (Zone 1–Zone 5) и модул „Сила“, изразени в часове/минути. Структурата позволява квантилен анализ и стратификация по нивото на тренираност (Q1–Q4).

Статистическата рамка включва описателен анализ (централна тенденция, вариация, асиметрия и ексцес), тестове за нормалност (Shapiro–Wilk), непараметрични тестове за зависими извадки (Wilcoxon), както и оценка на практическия ефект (размер на ефекта γ и мощност на теста). Тази рамка е използвана както за оценка на сходството между обучаващи и генерирани програми, така и за верификация на ефекта от експерименталното приложение.

Използван е и контрол на качеството на данните (валидиране на типове, диапазони, липсващи стойности и консистентност между променливите), както и стандартизация на входните и изходните параметри за осигуряване на стабилно обучение и ограничаване на риска от „mode collapse“ при генеративната архитектура.

5. ТЕОРЕТИЧНА РАМКА: ТРЕНИРОВЪЧНОТО НАТОВАРВАНЕ КАТО ЗАДАЧА ЗА ОПТИМИЗИРАНЕ

5.1. Понятиен апарат и структуриране на натоварването

В дисертационния труд тренировъчното натоварване се разглежда като многокомпонентна система, характеризирана чрез величина, характер и насоченост. Величината на натоварването се описва чрез обем и интензивност; характерът - чрез координационна сложност и степен на сходство със състезателната дейност; насочеността - чрез целевата функция на въздействието върху вегетативни, двигателни и психомоторни функции. Йерархичният принцип и дедуктивният подход позволяват натоварването да бъде разглеждано на различни нива, достигащи до количествено измерими параметри, необходими за формализиране и прецизно дозиране.

Централно място заема връзката „доза-ефект“ и диапазонът на въздействие, в който натоварването действа като формиращ стимул, предизвикващ положителни функционални и структурни промени. Критериите за дозиране определят различни категории натоварвания (максимални, средни, малки) и необходимостта от динамично равновесие между тях в тренировъчния процес.

5.2. Математическа постановка на оптимизационната задача

Тренировъчното натоварване се формализира като управляем обект, описан чрез вектор от управляващи параметри $x = (x_1, x_2, \dots, x_n)$, които могат да включват обем, интензивност, честота, интервали за възстановяване и др. Критерият за оптималност (целевата функция) се разглежда като $Q(x)$, а оптимизационната задача - като търсене на екстремум в допустимото пространство G_x при наличие на факторни и функционални ограничения.

В общ вид задачата се формулира като: намиране на $x^* \in G_x$, за което $Q(x^*) = Q_{\max}$ (или Q_{\min}), при спазване на ограничения от тип равенство и неравенство. Когато целевата функция и ограниченията са линейни, задачата се отнася към линейно програмиране; при нелинейни зависимости - към нелинейно програмиране. В контекста на тренировъчните натоварвания често се налага многокритериално оптимизиране и

използване на векторен критерий, поради едновременното управление на конкурентни цели (постижение, възстановяване, контрол на риска от травматизъм и др.).

5.3. Особености на целевите функции в тренировъчната практика

Целевите функции могат да бъдат унимодални или мултимодални, изпъкнали или вдлъбнати, както и оврагови и платовидни. Многоекстремалният характер на $Q(x)$ отразява наличието на множество локални екстремуми, обусловени от нелинейните физиологични зависимости и взаимодействията между управляващите параметри. Овраговостта се проявява при силно различаващи се размерности и чувствителност на параметрите, което затруднява сходимостта и надеждността на класическите локални методи.

5.4. Алгоритми за оптимизиране на сложни целеви функции

Дисертационният труд систематизира основни изисквания към алгоритмите за оптимизиране: сходимост, бърза сходимост, малка заемана оперативна памет, лесна подготовка на условията, надеждни критерии за спиране и изпълнение на ограниченията. В практиката липсва универсален алгоритъм, приложим към всички класове задачи, поради което изборът зависи от формата на целевата функция, броя параметри и ограниченията.

5.4.1. Метод на Гелфанд и Цетлин

Методът реализира оврагово търсене на екстремум при два или повече управляващи параметъра чрез локализация на екстремума по градиентно направление от две начални точки и последователно придвижване по линията на оврага. В контекста на цикличните спортове методът е интерпретиран чрез управление на параметри като обем, интензивност и функционални показатели (HR, LA и др.), при спазване на допустимите физиологични граници.

5.4.2. Метод на Nooke и Jeeves (метод на конфигурациите)

Методът е безградиентен подход за директно търсене, включващ базисно изследване и търсене по конфигурация. Основната хипотеза е локалната неизменност на направлението на търсене, което позволява ефективно придвижване по праволинейни оврази. В тренировъчния контекст се интерпретира като последователно изследване на параметри с $\pm \Delta x_i$ и избор на конфигурация, водеща до подобрение на целевата функция (например, показатели за умора).

5.4.3. Метод на Rosenbrock (въртяща се координатна система)

Методът използва ротация на координатната система така, че една от новите оси да съвпадне с направлението на най-бързо подобрение. Това улеснява локализацията на екстремума в завъртяното пространство. В спортната интерпретация методът подпомага търсенето на оптимални комбинации на взаимно свързани параметри (напр. честота и сила), когато изолиниите на целевата функция са наклонени спрямо първоначалните координатни оси.

5.4.4. Евристични и стохастични методи за глобално търсене

При многоекстремални функции се разглеждат евристични и стохастични подходи, включително методът на Price (кластерен анализ в популация от случайни точки) и методът на Wang и Luus (псевдоедномерно случайно търсене). Тези методи целят преодоляване на локални екстремуми чрез глобално изследване на параметричното пространство, като същевременно поставят високи изисквания към изчислителния ресурс при увеличаване на размерността.

5.4.5. Градиентни подходи за преодоляване на локални екстремуми

В дисертационния труд е разгледан методът на „тежкото топче“, който въвежда инерционен компонент (масата β) в рекурентната формула на градиентното движение, с цел прескачане на слабо изразени локални минимуми. Подобен механизъм може да подпомогне оптимизирането при мултимодални повърхнини, но не гарантира достигане на глобален екстремум от една начална точка.

5.4.6. Тунелен метод за глобално оптимизиране

Тунелният метод редува минимизираща фаза (локална оптимизация на $Q(x)$) и тунелна фаза, в която се дефинира помощна функция $T(x)$, насочваща търсенето към нови области на допустимото пространство, за да се избегнат локални екстремуми. В приложените примерни постановки методът илюстрира логика на последователно подобрене при спазване на ограничения.

Тунелният метод се разглежда като евристичен алгоритъм за глобално оптимизиране при многоекстремални целеви функции, при които стандартните локални процедури често попадат в локални екстремуми. Методът осигурява механизъм за контролирано напускане на областта на текущия локален минимум чрез конструиране на тунелна функция, която потиска вече намерените решения и насочва търсенето към нови области на допустимото пространство.

Алгоритъмът включва следните основни стъпки:

1. Избор на начална точка x_0 в допустимото пространство на решението.
2. Минимизираща фаза: намиране на локален минимум x^* на целевата функция $Q(x)$ чрез метод за локално търсене (градиентен или безградиентен), при спазване на ограниченията.
3. Формиране на тунелна функция $T(x)$, която зависи от $Q(x)$ и текущото най-добро решение x^* (съответно стойността $Q^* = Q(x^*)$).
4. Тунелна фаза: намиране на точка $x(T)$, при която $T(x(T)) = 0$ или $T(x(T)) \approx 0$ при числено решение, като целта е да се достигне област с по-добра стойност на $Q(x)$ спрямо Q^* .
5. Актуализация: приема се нова начална точка $x_0 = x(T)$ и алгоритъмът се повтаря от минимизиращата фаза.
6. Критерий за спиране: редуването на двете фази продължава до изпълнение на зададен критерий (например липса на подобрене или достигане на предварително зададена точност).

В дисертационния труд е представен примерен вариант на тунелна функция, при която $T(x) = Q(x) - Q^*$ за $Q(x) > Q^*$ и $T(x) = 0$ в противен случай. Тази постановка изисква в

тунелната фаза да се намери точка, за която $Q(x(T)) < Q^*$, което гарантира преминаване към по-добро решение и последващо локално дооптимизиране.

Теоретичната интерпретация на метода показва, че тунелната фаза играе ролята на механизъм за глобално претърсване, докато минимизиращата фаза осигурява прецизиране в локален смисъл. Тази логика е съвместима с оптимизационни задачи при наличие на ограничения и конкурентни критерии, характерни за моделирането на тренировъчни натоварвания.

Направеният теоретичен анализ обосновава необходимостта класическите оптимизационни методи да бъдат допълнени с генеративни модели на изкуствения интелект, които да осигурят глобален синтез и селекция на тренировъчни сценарии при голям брой степени на свобода.

В дисертационния труд тренировъчното натоварване е формулирано като задача за оптимизиране на управляващи параметри, които описват величина, характер и насоченост на въздействието във времето. Управляващият вектор включва компоненти за обем (време), интензивност (зонова структура), честота и редуване на натоварване и възстановяване, като се отчита позиционирането спрямо контролни и основни състезания.

Общата математическа постановка е представена чрез целева функция (или вектор от целеви функции) и система от ограничения, които отразяват физиологичната допустимост и методическите принципи. В скаларен вид задачата може да се представи като: $\min_u f(u)$ при условия $g_j(u) \leq 0$, $h_k(u) = 0$, $u \in U$, където U е допустимото пространство на управляващите параметри, а g_j и h_k задават ограничения, свързани с пределни обеми, интензивности, редуване на режими и календарни фиксации.

При многокритериално оптимизиране целта се задава като вектор $F(u) = (f_1(u), f_2(u), \dots, f_m(u))$, например: максимизиране на функционалния ефект, минимизиране на риск от преумора/травми, съхраняване на техника и координационна сложност, и минимизиране на отклонението от методическите норми. Решението се търси в клас Pareto-ефективни решения, от които се избира конкретна програма според приоритети и ограничения (Pareto-фронт).

Теоретичният анализ показва, че при реални тренировъчни програми целевите функции са сложни, нелинейни и често недиференцируеми, с множество локални екстремуми. Това обосновава използването на класически безградиентни и хибридни методи (Hooke-Jeeves, Rosenbrock, Price, тунелен метод и др.) за локално усъвършенстване, както и необходимостта от глобално изследване на пространството на решенията чрез генеративни модели, които синтезират кандидати със съхранена структурна логика.

Важен извод от теоретичната част е, че оптимизацията в спорта не може да бъде сведена до един показател или определен момент във времето. Затова се акцентира върху микроцикличната и мезоцикличната структура, върху последователността от натоварване и възстановяване и върху периодизацията, която задава контекст за интерпретация на всяко дневно натоварване като елемент от динамична система.

6. ОПТИМИЗИРАНЕ НА ТРЕНИРОВЪЧНОТО НАТОВАРВАНЕ ЧРЕЗ ГЕНЕРАТИВНИ МОДЕЛИ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

6.1. Теоретични основи и типове генеративни модели

Генеративните модели се разглеждат като алгоритми за моделиране на неизвестно вероятностно разпределение $p(x)$ на наблюдаваните данни и последващо генериране на нови реалистични реализации $x_{\text{нов}}$. В дисертационния труд е направен сравнителен анализ на основните архитектури: вариационни автоенкодери (VAE), генериращи състезателни мрежи (GAN), дифузионни модели и normalizing flows, като са разгледани техните принципи на обучение и приложимост към времеви редове и структурирани данни в спорта.

VAE интегрират енкодер-декодер структура с вероятностна интерпретация чрез ELBO, докато GAN реализират съревнователен механизъм между генератор и дискриминатор, водещ до синтез на данни с висока реалистичност. Дифузионните модели използват стъпков процес на добавяне и премахване на шум, а normalizing flows - обратими трансформации, позволяващи директно изчисляване на плътности.

Вариационните автоенкодери реализират компресиране на входните данни в латентно пространство z и последващо възстановяване чрез декодиране, като по този начин се моделират сложни зависимости при редуцирана размерност. Вероятностната постановка позволява да се работи с разпределения на латентните променливи и да се генерират нови реализации чрез пробирание и декодиране.

При генериращите състезателни мрежи обучението се основава на съревнователна динамика, при която генераторът се стреми да синтезира реализации, неразличими от реалните данни, а дискриминаторът оценява дали дадена реализация е реална или генерирана. В дисертационния труд се подчертава, че този механизъм може да доведе до висока „реалистичност“ на синтеза, но изисква внимателна настройка и контрол върху стабилността на обучението.

Flow-based моделите (normalizing flows) изграждат сложни вероятностни разпределения чрез последователност от обратими преобразувания на базово разпределение. Обратимостта позволява директно изчисляване на плътности и оценка на вероятността на всяка генерирана реализация, което е особено полезно при задачи, в които е необходим количествен контрол върху „правдоподобността“ на предложените решения.

Дифузионните модели се описват като процес на постепенно въвеждане на шум в данните и последващо възстановяване на първоначалните структури чрез обратен (денойзинг) процес, реализиран на стъпки. Този принцип поддържа стабилност на обучението и е приложим не само за изображения, но и за структурирани данни и времеви редове, когато се цели възпроизвеждане на сложни емпирични закономерности.

В дисертационния труд е разгледана и условната постановка на генеративните модели, при която те се управляват от контекстни променливи. В тренировъчния контекст това позволява да се моделира условно разпределение $p(x|y)$, където y включва физиологични и календарни входове (например $VO_{2\text{max}}$ и оставащи дни до контролно или основно състезание), а x представя структурата на дневното натоварване по зони и силова подготовка.

По този начин тренировъчната програма може да бъде интерпретирана като структуриран времеви ред, в който всеки дневен запис се описва чрез вектор от натоварвания по зони на интензивност и силова компонента. Генеративните модели възпроизвеждат закономерности на периодизацията, повторемостта на микроцикли и ограниченията за допустими комбинации, което създава основа за интегрално моделиране и оптимизиране.

6.2. Предимства и ограничения на генеративния подход в тренировъчния контекст

Основно предимство на генеративните модели е способността им да генерират множество алтернативни тренировъчни сценарии, които удовлетворяват факторни и областни ограничения, като по този начин подпомагат треньорските решения в условията на многообразие от степени на свобода. Генеративният подход позволява виртуално „експериментиране“ в параметричното пространство, включително изследване на редки или екстремни комбинации без непосредствен риск за спортистите.

Ограниченията са свързани с качеството и обхвата на данните, изчислителната сложност на обучението, рискове като *mode collapse* при GAN и трудности при интерпретацията на моделите като „черна кутия“. Това налага строг механизъм за верификация и експертна оценка, както и интеграция на логически правила и методически филтри.

В дисертационния труд генеративният етап се разглежда като компонент за глобално претърсване, който може да бъде комбиниран с последващо локално усъвършенстване и контрол. Обобщена е последователност от стъпки, типична за „Global-Local Synergy“, при която се преминава от обучение на генеративен модел и мащабно пробиране на вероятностното пространство, към локална донастройка в посока „обем-интензивност“ и периодична верификация на решенията в рамките на микроциклъ.

Като актуално направление е разгледан подходът „Tunnel Generative + Adaptive Search“, който съчетава обучение на GAN, „тунелно“ прогресивно стесняване на набора от кандидат-решения (*tunnel layers*) и адаптивно търсене чрез непрекъсната актуализация на модела със сигнали от реалното приложение. Тази логика подчертава необходимостта генерирането да бъде не еднократен акт, а процес с обратна връзка и динамична проверка на приложимостта на сценарии при променящи се условия.

В тренировъчния контекст това означава, че генеративният модел следва да бъде интегриран със система за логическа и физиологична верификация, която да контролира вътрешната съгласуваност на програмата, съвместимостта между зони и силова работа и допустимите серии от последователни високоинтензивни дни. По този начин се минимизира рискът от генериране на методически недопустими натоварвания при запазване на статистическата близост до обучаващите данни.

Таблица 1. Сравнение на основни класове генеративни модели, разгледани в дисертационния труд.

Клас модел	Ключов механизъм	Силни страни в тренировъчния контекст	Ограничения и рискове
VAE	Енкодер-декодер с латентно пространство z и вероятностна постановка (ELBO).	Устойчиво обучение при структурирани данни; възможност за условна генерация и контрол чрез латентните фактори; удобно пробиране на допустими сценарии.	Възможно „усредняване“ на редки режими; необходима е параметризация на латентното пространство и контрол на реконструкцията.
GAN	Състезателен процес генератор-дискриминатор за синтез на реализации.	Висока реалистичност; ефективен за сложни разпределения и генериране на разнообразни кандидат-решения за планиране на натоварването.	Риск от нестабилност и mode collapse; чувствителност към голям брой параметри; нужда от регулиране и строг мониторинг.
Дифузионни модели	Стъпков процес на добавяне на шум и обратен денойзинг за възстановяване на структурите.	Стабилен принцип на обучение; възпроизвеждане на сложни закономерности; приложимост към времеви редове и структурирани данни.	Висока изчислителна сложност; повече стъпки при генериране; зависимост от настройката на процеса и критерия за спиране.
Normalizing flows	Поредица от обратими трансформации на базово разпределение с директна оценка на плътности.	Точна оценка на вероятността на генерираните реализации; контрол върху правдоподобност и сравнение на разпределенията.	Ограничения от изискването за обратимост; архитектурни компромиси при силно нелинейни зависимости и висока размерност.

Анализът показва, че изборът на архитектура следва да бъде съобразен с целта на модела (структурно сходство на разпределенията, разнообразие на кандидат-решенията, възпроизводимост на микроциклични закономерности) и с наличния обем и качество на данните. Независимо от избора, прилагането в спортната практика изисква интеграция с критерии за оценка и верификация, гарантиращи методическа допустимост и физиологична съвместимост на генерираните натоварвания.

6.3. Съвместимост на многокритериалните алгоритми с генеративните модели

В дисертационния труд оптимизацията се разглежда като многокритериална задача, при която се балансират конкурентни цели (спортно постижение, възстановяване, контрол на риска от травматизъм, рационално използване на времето). Генеративните модели се използват за създаване на богата начална популация от кандидат-решения, които впоследствие се оценяват и усъвършенстват чрез многокритериални алгоритми и Pareto-ориентиран анализ.

Предложената логика „генериране - оценка - оптимизация“ може да бъде реализирана итеративно, при което новопостъпилите данни от системи за мониторинг и контролни тестове допълнително усъвършенстват модела и намаляват неопределеността.

6.4. Интеграция на генеративните модели с тренировъчните програми: хибриден подход

Хибридният подход комбинира глобалните възможности на генеративните модели (VAE/GAN) за изследване и синтез на допустимото пространство с традиционни локални методи за оптимизиране, които осигуряват фина настройка на параметрите. Така се намалява рискът локалните методи да стартират от нерелевантни начални точки и да попаднат в локални екстремуми.

Ключов елемент е въвеждането на контекстни и условни променливи (например етап от годишния цикъл, оставащи дни до контролно/основно състезание), чрез които се осигурява методическа допустимост и съответствие с логиката на периодизацията.

В дисертационния труд са разгледани основни класове генеративни модели (VAE, GAN и дифузионни модели), като е обосновано, че при задачи за генериране на тренировъчни програми ключово изискване е възпроизводимостта на разпределенията по зони на натоварване и запазване на вътрешната времева логика. Това поставя акцент върху условното (conditional) генериране и върху контрола на несъответствия чрез допълнителни критерии и правила.

Обсъдени са характерните проблеми при генеративните архитектури, включително нестабилност на обучението и „mode collapse“ при GAN, което може да доведе до ограничено разнообразие и методически непълноценно синтезиране. В методиката са заложили процедури за контрол на входните данни, стратификация по VO_{2max} , синхронизация генератор–дискриминатор и периодично валидиране върху отделен набор, което ограничава риска от свръхадаптация и повишава устойчивостта на обучението.

Представени са и хибридни алгоритми, описани в научната литература („Global–Local Synergy“, „Tunnel Generative + Adaptive Search“, „Multi-Objective Generative Pipeline“), при които генеративното ядро създава набор от кандидати, а локални оптимизационни процедури извършват фина настройка спрямо конкретни цели и ограничения. В труда тези подходи са анализирани критично и са използвани като основа за изграждане на цялостна интегрална методика с ясно дефинирани етапи и критерии за верификация.

Разгледани са и стратегии за обучение при ограничени данни, което позволява адаптиране на модела към специфични профили, без загуба на зависимости, усвоени от по-голяма обучаваща извадка.

7. ИНТЕГРАЛНА МЕТОДИКА ЗА ОПТИМИЗИРАНЕ НА ТРЕНИРОВЪЧНОТО НАТОВАРВАНЕ ЧРЕЗ ГЕНЕРАТИВНИ МОДЕЛИ НА ИЗКУСТВЕН ИНТЕЛЕКТ

Под интегрална методика се разбира система от целенасочени средства, методи и форми за оптимизиране на тренировъчното натоварване в цикличните спортове на базата на утвърдени принципи посредством генеративни модели на изкуствения интелект. Методиката е линейна по вид и структурирана в строга логическа последователност, като всеки етап следва завършването на дейностите по предходния.

7.1. Етап 1: Проектиране на информационната и функционалната архитектура

- Дефиниране на целите на генеративния модел и описание на очакваните резултати чрез измерими метрики.
- Проектиране на масиви от данни за обучение: планиране на обем и обхват, схема на базата данни, релационни връзки, индексирание, метаданни и контрол на версии.
- Проектиране на изходните резултати: структура на персонализираните тренировъчни програми, формат на представяне, интерфейсни и визуализационни решения.
- Контекстно описание: семантика на данните, зависимости между променливите, сценарии за използване, динамика във времето.
- Избор на методи и средства за събиране на данни (вкл. експертно генерирани данни) и определяне рамките на обучаващата и тестовата извадка.
- Проектиране на тестови сценарии за работа с модела, включително стандартни и екстремни условия.

Етап 1 обхваща дейности по целеполагане, проектиране на извадки и масиви от данни, описание на метаданни, избор на методология за събиране на данни и избор на тестови сценарии, като целта е да се осигури информационна и функционална готовност за обучение, оценка и приложение на генеративния модел.

7.1.1. Дефиниране на целите на генеративния модел

Подетапът е насочен към формулиране на конкретни цели и очаквани резултати от приложението на генеративния модел за оптимизиране на тренировъчното натоварване. Целеполагането следва да бъде насочено към възможностите на модела да генерира персонализираните тренировъчни програми на базата на утвърдени принципи, като резултатите се представят чрез ясно дефинирани и измерими метрики, осигуряващи обективна оценка на ефективността на програмите и техния функционален ефект.

7.1.2. Проектиране на масиви от данни за обучение на модела

Проектирането на масивите от данни е свързано с планиране на обема и обхвата на данните, подготовката и структурирането им по начин, който максимизира тяхната полезност и ефективност в процеса на обучение.

- Планиране на обема и обхвата на данните: определяне на необходимия обем данни за обучение и на обхвата на данните, които ще бъдат включени в масива, включително времеви диапазон и релевантни параметри.
- Създаване на схема на базата данни: дефиниране на релационните връзки между различните таблици и набори от данни и избор на подходящи индекси за ускоряване на достъпа и обработката.
- Дефиниране на метаданни: създаване на документация за всяко поле (значение, тип, допустими стойности и връзки) и контрол на версиите на данните и схемата с цел проследимост и възстановяване.
- Интеграция и съвместимост на данните: стандартизация на форматиране и кодиране и разработване на процедури за обединяване на данни от различни източници в единен масив при гарантирана консистентност и целостта.

7.1.3. Проектиране на изходните резултати

Възможностите на модела за създаване на персонализирани тренировъчни програми налагат ясна структура и прецизност в дизайна на изходните резултати, както и дефиниране на измерими метрики за оценка на ефективността.

- Дефиниране на персонализирани тренировъчни програми: създаване на програми, които са съобразени с индивидуални цели, физически способности и текущо състояние на спортиста.
- Интерфейс за представяне на резултатите: разработване на информативни визуализации на резултатите (текст, графики, таблици, интерактивни панели) с цел подпомагане на експертната интерпретация и контрол.

7.1.4. Контекстно описание на данните

Контекстното описание на данните осигурява семантичното разбиране и правилна обработка на входните параметри и намалява вероятността от генериране на нерелевантни изходни резултати.

- Семантика на данните: включване на контекстуална информация за съдържателното значение на данните, дефиниции на измервателни единици, описания на категории данни и взаимоотношения между променливи.
- Връзки и зависимости: описание на контекстуални зависимости между данните, включително влияние на фактори върху тренировъчния процес и взаимовръзки между показатели.
- Примери и сценарии: предоставяне на примери за реални данни и сценарии на използване за подпомагане на интерпретацията в различни ситуации.
- Динамика на данните: описание на времеви аспекти (честота на актуализация, времеви зависимости) и динамични връзки между променливите.

7.1.5. Избор на методи и средства за събиране на данни

Подетапът включва избор на методи и средства за събиране на масиви от данни, необходими за обучение на генеративния модел, като методите се класифицират като придобиване, събиране, извличане и трансфериране на данни.

- Придобиване на данни: уеб скрапинг; API интеграции за достъп и извличане на данни от външни системи; сензори и IoT устройства за събиране на данни в реално време.
- Събиране на данни: анкети и въпросници; наблюдение; фокус групи и интервюта.
- Извличане на данни: data mining за извличане на скрити модели и тенденции; text mining за анализ на текстови данни; обработка на големи обеми от данни (Big Data); експертно генерирани данни, които съдържат логически последователни и валидирани структури за обучение.
- Трансфериране на данни: файлов трансфер (FTP/SFTP), облачно съхранение, репликация на бази данни, потоци от данни (message queues и data streams) за обмен в реално време.

7.1.6. Определяне на рамките на обучаващата и тестовата извадка

Подетапът е свързан пряко с обучението и тестването на модела на последващо йерархично ниво от методологията и включва дефиниране на обучаващата и тестовата извадка.

- Дефиниране на обучаващата извадка: обхват на данните и разнообразие в обучаващите данни; репрезентативност на извадката за генералната съвкупност, за която моделът се отнася; анализ за достатъчност за определяне на минималния обем данни, осигуряващ адекватна точност.
- Дефиниране на тестовата извадка: независимост от обучаващите данни за обективна оценка; репрезентативност; оптимален обем и статистическа значимост на резултатите.

7.1.7. Проектиране на тестови сценарии за работа с генеративния модел

Проектирането на тестови сценарии е съществен подетап за оценка на точността, надеждността, адаптивността и ефективността на модела при реални условия, като сценариите следва да обхващат широк спектър от ситуации.

- Дефиниране на целите на тестовите сценарии: определяне на целите за оценка на точността и параметризиране на модела при различни сценарии.
- Разработване на тестови сценарии: сценарии за спортисти с различни нива на подготовка; стандартни тренировъчни условия, отразяващи ежедневната практика; екстремни тренировъчни условия за проверка на устойчивостта.

Разработването на тестовите сценарии има за цел да осигури емпирично обоснована рамка за проверка на способността на генеративния модел (GAN) да възпроизвежда реалистични тренировъчни програми, отразяващи закономерностите, установени в обучаващия масив от данни. В този смисъл контролът на сценариите представлява едновременно инструмент за параметризация (чрез задаване на входни условия) и инструмент за верификация (чрез проверка на очакваните изходни зависимости).

Тестовите сценарии са проектирани така, че да отразяват различни нива на аеробен капацитет (VO_{2max}) и различна времева отдалеченост до контролни и основни състезания. В сценарния набор са включени стойности на VO_{2max} в диапазона 50–80 ml/kg/min, което позволява стратифициране на състезателите по четири кватилни подгрупи (Q1–Q4) – от ниска до висока тренираност.

Времевата отдалеченост до контролни състезания (Control_days) служи за проверка дали моделът възпроизвежда закономерностите в дните непосредствено преди контролно състезание ($t-1$), както следва:

- липса на високоинтензивни режими (Zone 4–Zone 5);
- отсъствие на силови натоварвания (Strength = 0);
- наличие на компенсаторна нискоинтензивна работа в Zone 1.

Времевата отдалеченост до основни състезания (Main_days) в анализираниите програми варира между 81 и 321 дни, което дефинира дълъг и много дълъг период до основен старт. Тази променлива задава сезонния контекст на натоварването и подпомага разграничаването между базова подготовка и етапи, в които се очаква по-голям относителен дял на специфична работа.

На базата на комбинациите между VO_{2max_input} , Control_days и Main_days са дефинирани пет основни тестови сценария (S1–S5), които симулират различни типове тренировъчни ситуации и налагат конкретни методически ограничения върху изхода на генеративния модел:

Сценарий 1 (S1) – нисък VO_{2max} (Q1), Control_days = 1, Main_days \geq 12 седмици. Сценарий за слабо тренирани биатлонисти, при който се очаква пълно отсъствие на високоинтензивни натоварвания (Zone 4–5), липса на силова работа и кратък компенсаторен обем в Zone 1.

Сценарий 2 (S2) – среден VO_{2max} (Q2–Q3), Control_days = 1, Main_days \geq 12 седмици. Сценарий за състезатели със средно ниво на подготовка, характеризиращ се с умерен обем в Zone 1 при отсъствие на Zone 4–5 и Strength.

Сценарий 3 (S3) – висок VO_{2max} (Q4), Control_days = 1, Main_days \geq 12 седмици. Сценарий за високо тренирани биатлонисти, при който се очакват нулеви стойности в Zone 4–5 и Strength с наличие на кратка вработваща тренировка в Zone 1.

Сценарий 4 (S4) – нисък VO_{2max} (Q1), Control_days = 1, Main_days \gg 12 седмици. Сценарий за базов подготвителен етап при слабо тренирани биатлонисти, при който ден $t-1$ е възстановителен или с аеробна работа с ниска интензивност и без силови натоварвания.

Сценарий 5 (S5) – висок VO_{2max} (Q4), Control_days = 1, Main_days \gg 12 седмици. Сценарий за подготвителен етап при високо тренирани биатлонисти, при който се очаква контролирано редуциране на интензивността преди контролното състезание при запазване на нискоинтензивна вработваща компонента.

7.2. Етап 2: Агрегиране на масиви от данни за обучение

Етапът включва идентифициране на източници, типове данни и достъп до тях, както и събиране и обединяване на данни в единен масив. Акцент се поставя върху структурирани таблични данни, стандартизация на единици и формати и контрол на консистентността.

В методиката Етап 2 се конкретизира чрез подетапи, насочени към осигуряване на релевантност, репрезентативност и технологична възпроизводимост на данните за обучение на генеративния модел.

- Идентифициране на източниците, типа на данните и достъпа до тях: избор на релевантни източници (вътрешни бази данни, публични хранилища, приложни програмни интерфейси и др.) и класификация на данните като структурирани и неструктурирани.
- Извличане на данни: извличане на полезна информация от предварително събрани данни, включително data mining за извличане на модели и тенденции в големи масиви и експертно генерирани данни, които съдържат логически последователни, валидирани структури за обучение на генеративни модели.
- Трансфериране на данни: предаване на данни между системи (файлов трансфер, облачно съхранение, репликация, потоци от данни), с цел обединяване в единен масив и гарантиране на целостта и проследимостта на промените.
- Определяне на рамките на обучаващата и тестовата извадка: дефиниране на обхват, репрезентативност и достатъчност на обучаващата извадка, както и независимост и оптимален обем на тестовата извадка за обективна оценка на производителността.
- Проектиране на тестови сценарии: разработване на сценарии за проверка на точността, надеждността, адаптивността и ефективността на модела при различни контексти и условия на тренировъчния процес.

7.3. Етап 3: Обработка на масивите от данни

- Изчистване на данните от грешки: идентифициране и коригиране на аномалии, премахване на дублиращи записи, обработка на липсващи стойности, стандартизиране на формати.
- Валидиране на данните спрямо предварително дефинирани диапазони и типове; проверка на обхвата и ограниченията.
- Съхранение на данните: избор на инфраструктура (SQL/NoSQL/облак), защита (контрол на достъп, криптиране), архивиране и възстановяване, проследимост.

7.4. Етап 4: Управление на генеративния модел

- Контрол на входящите и изходящите данни чрез специфични правила и процедури за валидация.
- Обучение на генеративния модел (обучение с един/няколко примера; фина настройка), съобразено със спецификата на спортната дисциплина и наличните данни.
- Настройка на мрежата за обучение: скорост на обучение, брой епохи, валидиране по време на обучение; метрики (CAГ, KL-дивергенция, дял на логически несъответствия).
- Синхронизация на обучението (генератор - дискриминатор) и управление на рисковете като mode collapse при GAN.

7.5. Етап 5: Оценка на ефективността на модела

- Определяне на критерии за успех, минимални приемливи прагове и механизъм за експертна оценка.
- Метрики за разлика между реални и генерирани стойности (напр. средна абсолютна грешка).
- Мерки за сходство на разпределенията (KL-дивергенция) и анализ на структурни зависимости.
- Проверка на логическа и физиологична съгласуваност чрез формализирани правила.
- Емпирично валидиране чрез контролни експерименти и обратна връзка.

7.5.1. Формализирани логически цели и правила за параметризация и верификация

В рамките на методиката са дефинирани цели и правила, извлечени от емпиричния анализ на обучаващите програми, които служат едновременно като условия за параметризация на GAN и като критерии за логическа верификация на генерираните резултати:

1. Да се гарантира, че входните стойности на VO_2max попадат в интервала $[50; 80] \text{ ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$ и следват реалното разпределение с квантилни граници $Q1=56$, $Q2=62$, $Q3=71$.
2. Да се осигури стратифицирано генериране по VO_2max ($Q1-Q4$) за намаляване на вътрешната дисперсия и по-прецизно моделиране на зависимостите между VO_2max и натоварването.
3. Да се възпроизведе монотонната зависимост между VO_2max и обема на натоварване, като за по-високи квантилни подгрупи ($Q3-Q4$) се генерират по-високи медианни стойности на дневния обем и на зоните Zone 1–3.
4. Да се запази аеробната доминанта, при която Zone 1–2 формират $\geq 75\%$ от общия дневен обем, а зоните Zone 4–5 и Сила са ограничени до $\leq 25\%$.
5. Да се възпроизведат реалните дялове на нулеви стойности при високоинтензивните и силовите режими в интервали, съвместими с обучаващата извадка (Zone 4; Zone 5; Сила).
6. Да се съхрани честотата на високоинтензивни дни при най-високия VO_2max квантил ($Q4$), така че приблизително половината от дните да съдържат интензивен компонент ($Z4+Z5>0.1$).
7. Да се спазва принципът на взаимно изключване между високоинтензивни и силови натоварвания, така че в $\geq 85\%$ от дните с ($Z4+Z5>0.1$) да е изпълнено условието Сила = 0.
8. Да се предотвратява натрупване на прекомерна интензивност, като не се допуска повече от два последователни дни с ($Z4+Z5>0.1$), както и ден с висока интензивност, следван непосредствено от ден със силова подготовка.
9. Да се гарантира вътрешна логическа съгласуваност на микроцикъла, като се запазва последователност от натоварване и възстановяване, характерна за реалните тренировъчни програми.

8. ЕМПИРИЧНА БАЗА, ДАННИ И ПАРАМЕТРИЗАЦИЯ НА ГЕНЕРАТИВНИЯ МОДЕЛ

Обучението и тестването на генеративния модел се основава на експертно разработени тренировъчни програми за младежкия национален отбор по биатлон, структурирани на дневно ниво. Достъпът до тренировъчните програми е реализиран чрез файлов формат Excel (.xlsx) като структуриран и машинночетим масив, включващ 44 694 наблюдения (дневни тренировъчни единици) и 200 еднозначно идентифицирани тренировъчни програми (Program_ID).

8.1. Входни и изходни параметри

Входните параметри, използвани за персонализиране на генерираните тренировъчни програми, са:

- VO₂max_input - максимална кислородна консумация (ml/kg/min), използвана за индивидуализиране на натоварването според аеробния капацитет;
- Control_days - брой дни до следващо контролно състезание;
- Main_days - брой дни до следващо основно състезание.

Изходът на модела е дневно разпределение на тренировъчния обем по зони на натоварване (Zone 1- Zone 5) и силова подготовка (Strength), изразено в часове, както и тип на деня (Type), обозначаващ тренировъчен ден или ден на контролно/основно състезание.

8.1.1. Описание на променливите (структура на дневния запис)

В Таблица 2 е представена нотацията на наблюденията, използвана в обучаващата и тестовата извадка. Тя обединява профил на състезателя, календарен контекст и разпределение на дневния тренировъчен обем по зони на натоварване.

Всеки дневен запис в масива от тренировъчни програми съдържа комбинация от количествени и качествени показатели, които описват: (1) физиологичен профил на състезателя (VO₂max_input), (2) параметри, свързани с участието в състезания (Control_days, Main_days, Control_dates, Main_dates), (3) хронологичното позициониране на тренировката (Day_index) и (4) детайлното разпределение на тренировъчния обем по зони на натоварване (Zone 1–Zone 5) и силова подготовка (Strength), както и (5) тип на деня (Type), обозначаващ провеждане на контролно или основно състезание.

Нотацията на основните променливи в дневния запис може да бъде обобщена както следва:

- Program_ID – уникален идентификатор на тренировъчна програма, използван за групиране на дневните записи в рамките на една и съща програма;
- VO₂max_input – максимална кислородна консумация (ml/kg/min) като входен параметър за индивидуализиране на натоварването според аеробния капацитет;
- Control_days / Main_days – оставащи дни до контролно/основно състезание като контекстни (условни) променливи;
- Control_dates / Main_dates – дати на контролни и основни състезания като календарен маркер на състезателния контекст;

- Day_index – пореден номер на деня в седмицата (1 = понеделник, ..., 7 = неделя), използван за хронологично подреждане и микроцикличен анализ;
- Zone 1–Zone 5 – продължителност (в часове) на работа в съответната зона на натоварване, от нискоинтензивна аеробна работа (Zone 1) до максимална интензивност (Zone 5);
- Strength – време (в часове), отделено за силова подготовка;
- Type – обозначение за провеждане на контролни и основни състезания.

С цел по-ефективно моделиране на взаимовръзките между основните компоненти на наблюденията и за осигуряване на проследимост на ниво биатлонист – програма – ден – натоварване е приложена релационна структура на данните, организирана в логически таблици (профил, програма, дневна сесия и натоварване). Тази организация осигурява нормализация на данните, мащабируемост на системата и възможност за индексирание по ключови полета (напр. Program_ID и Day_index), което улеснява обучението и валидацията на генеративния модел и изпълнението на условни запитвания към подмножества от данни.

Таблица 2. Описание на променливите в дневните тренировъчни наблюдения.

Променлива	Описание	Тип стойности
Program_ID	Уникален идентификатор на тренировъчна програма; служи за групиране на дневни записи към една и съща програма.	Текст
VO2max_input	Входен параметър – максимална кислородна консумация ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$), използвана за индивидуализиране на натоварването според аеробния капацитет.	Цяло число
Control_days	Брой дни до контролно състезание; използва се за мониторинг на формата и адаптацията.	Цяло число
Main_days	Брой дни до основно състезание; задава целеви пик на спортната форма.	Цяло число
Control_dates	Дати за провеждане на контролни състезания/тестове (напр. 10 km бягане, стрелкови тестове).	Дата/Текст
Main_dates	Дати за провеждане на основни състезания.	Дата/Текст
Day_index	Пореден номер на деня в седмицата (1=Понеделник ... 7=Неделя) за хронологично подреждане	Цяло число

	и анализ на микроцикъла.	
Zone 1	Продължителност на тренировки в зона 1 (нискоинтензивна аеробна работа/възстановителни сесии), в часове.	Десетично число
Zone 2	Продължителност на тренировки в зона 2 (умерена аеробна работа – базова издръжливост), в часове.	Десетично число
Zone 3	Продължителност на тренировки в зона 3 (смесена работа – преход аеробно/анаеробно), в часове.	Десетично число
Zone 4	Продължителност на тренировки в зона 4 (висока интензивност около анаеробния праг), в часове.	Десетично число
Zone 5	Продължителност на тренировки в зона 5 (максимална интензивност; кратки интервали/симулации), в часове.	Десетично число
Strength	Време за силова подготовка (сила, експлозивност, стабилност), в часове.	Десетично число
Type	Тип на деня: тренировъчен ден / контролно състезание / основно състезание.	Текст

8.2. Стратификация по VO_2max и медианни дневни натоварвания

За намаляване на вътрешната вариация и по-прецизно моделиране на зависимостите между аеробния капацитет и тренировъчните натоварвания е приложена стратификация по кватилни подгрупи на VO_2max (Q1-Q4). Таблица 3 представя медианни дневни натоварвания (в часове) по зони на натоварване и общ обем.

Вариационният анализ на входната променлива VO_2max показва широк размах (мин. = 50; макс. = 80 ml/kg/min) и умерена положителна асиметрия ($Sk \approx 0,29$), което означава концентрация на наблюденията в по-ниски стойности и наличие на дясна опашка с порядко срещани високи стойности. Отрицателният ексцес ($K \approx -1,133$) свидетелства за по-плоска форма на разпределението и по-тънки опашки спрямо нормалното разпределение. Deskриптивните показатели за VO_2max се обобщават чрез квантилни оценки: Q1 = 56; медиана = 62; Q3 = 71 ml/kg/min, при средна аритметична стойност 63,69 и популационно стандартно отклонение $\sigma = 8,736$. Този профил подкрепя използването на

непараметрични оценки като устойчиви измерители на вариацията и аргументира необходимостта от стратификация по квантилни групи (Q1–Q4).

Стратификацията ограничава вътрешната дисперсия в отделните подгрупи и позволява условна параметризация на генеративния модел, при която зависимостите между аеробния капацитет и тренировъчните натоварвания се моделират по-точно за всяко ниво на тренираност. На практика това означава, че генерираните дневни натоварвания се модулират спрямо принадлежността към Q1–Q4, като се запазва емпирично установената тенденция за по-високи медианни стойности на Zone 1–Zone 3 и на общия обем при по-високи нива на VO_{2max} .

Таблица 3. Квантилно разпределение на медианни дневни натоварвания (по VO_{2max})

VO_{2max} квантил	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Сила	Общо натоварване
Q1	43.9	16.1	3.2	0	0	0	75.7
Q2	46.8	17.8	3.2	0	0	0	80.7
Q3	52.3	19.6	3.4	0	0	0	89.4
Q4	59.3	20.8	3.7	0	0	0	93.8

Медианните стойности показват ясна монотонна зависимост между нивото на аеробния капацитет и обема на натоварването: с нарастването на VO_{2max} се отчита постепенно увеличение на Zone 1-3 и на общия обем. Високите зони (Zone 4-5) и силовите натоварвания са със стойности, близки до нула, което отразява доминирането на ниско- и умереноинтензивна работа при базова подготовка.

Допълнителна параметризация е реализирана чрез индикаторни зависимости (условни вероятности), които описват честотата на съвместни прояви на висока интензивност и силова подготовка в реалните програми и се използват като контролни ограничения при генерирането.

- $P(Z4+Z5>0.1 \mid VO_{2max} \in Q4) = 0,5087$ – вероятността сумарното натоварване в Zone 4 и Zone 5 да надвишава 0,1 ч (≈ 6 минути) при условие, че състезателят е в най-високия квантил на VO_{2max} (Q4). Стойността показва, че приблизително в 51% от дните за тази група има отчетлива високоинтензивна компонента.

- $P(Сила = 0 \mid Z4+Z5>0.1) = 0,8547$ – вероятността силовото натоварване да бъде нула при наличие на високоинтензивно натоварване ($Z4+Z5>0.1$). Високата стойност означава, че при около 85% от случаите, когато е налице висока интензивност, отсъства силова тренировка в същия ден.

Тези зависимости отразяват принципа на функционална несъвместимост между силови и високоинтензивни натоварвания в рамките на микроцикъла и служат като логическа рамка за верификация на генерираните програми.

8.3. Достъп до данните и технологична реализация

В съответствие с методиката достъпът до тренировъчните програми е реализиран чрез файлов формат Excel (.xlsx), който представлява структуриран и машинно четим масив, подходящ за съхранение на таблични данни с хетерогенен тип (числови и текстови стойности).

Конкретната структура на файла включва: структурирана таблица в редове и колони, 44 694 наблюдения (дневни тренировъчни единици), еднозначно

идентифицирани програми (Program_ID) и структурирани променливи, които описват профила на състезателя, календарния контекст и разпределението на натоварването по зони на натоварване (Zone 1- Zone 5) и модул „Сила“ (Strength).

За целите на анализа и интеграцията в моделната архитектура данните са обработени чрез скриптова обработка с помощта на програмния език Python, посредством библиотеката pandas, която предоставя функционалност за директно зареждане на .xlsx файлове, конвертиране на данните в tabular (DataFrame) формат и последваща трансформация в структурни входове за модела.

Технологичната реализация е съобразена с изискванията за проследимост и възпроизводимост на обработките, като всяка трансформация (валидиране на типове, диапазони и консистентност, стандартизиране на формати и единици, агрегиране по дни и програми) се изпълнява по формализирани правила.

8.4. Правна съвместимост и етична обоснованост

Методиката разглежда в интегриран вид правния и лицензионен статус на използваните данни, файлови формати и софтуерни инструменти, като се цели гарантиране на законосъобразност, прозрачност и възпроизводимост на изследването.

В този контекст данните са използвани като структурирани тренировъчни програми, предоставени за научни цели, а обработката и анализът са реализирани чрез широко използвани софтуерни средства и формати, които осигуряват стандартизираност и контрол на достъпа.

9. ПРОЕКТ НА ГЕНЕРАТИВНИЯ МОДЕЛ И ВЕРИФИКАЦИЯ

9.1. Избор на архитектура и функционална постановка

В дисертационния труд е обоснован изборът на GAN архитектура като средство за генериране на тренировъчни програми. GAN моделът се разглежда като инструмент за глобален синтез на тренировъчни сценарии, при който генераторът създава нови дневни натоварвания, а дискриминаторът отличава реални от генерирани програми. Състезателният механизъм подпомага възпроизвеждането на статистическата структура на обучаващите данни.

9.2. Контрол и стандартизация на входните и изходните данни

Преди обучение е реализиран контрол на входните данни чрез валидиране на типове, диапазони и консистентност между променливите. Стандартизацията цели да ограничи влиянието на различните мащаби на параметрите и да подпомогне сходимостта на обучението. Контролът на изходните данни включва проверка за спазване на логиката на периодизацията, редуване между натоварване и възстановяване и съответствие с контекстните условия (Control_days, Main_days).

9.3. Метрики за оценка на качеството на генерираните програми

Оценката на качеството на генерираните тренировъчни натоварвания е проведена чрез: (1) метрики за разлика между реални и генерирани стойности (средна абсолютна

грешка - САГ), (2) мерки за сходство на разпределенията (KL-дивергенция), и (3) проверка на логическа и физиологична съгласуваност чрез формализирани правила.

Средната абсолютна грешка (САГ) се използва като индикатор за точността на модела при възпроизвеждане на дневните стойности (в часове/минути) по енергийни режими. За всеки режим на натоварване (Zone 1–Zone 5 и Strength) се изчислява средната абсолютна разлика между реалните и генерираните стойности на дневна база.

Сходството на разпределенията между реалните и генерираните програми се оценява чрез KL-дивергенция (Kullback–Leibler), изчислена върху нормализирани относителни дялове по зони на натоварване:

Количествените метрики (САГ и KL-дивергенция) са допълнени с формализирани логически правила за контрол на вътрешната съгласуваност и методическата допустимост (например ограничения върху натоварванията в дни, свързани с контролни и основни състезания, и ограничения върху последователността на високоинтензивни дни в микроцикъл).

9.4. Анализ на структурното сходство между обучаващи и генерирани програми

Чрез анализ на микроцикличната структура е установено, че приблизително три четвърти от програмите в обучаващата и генерираната извадка съдържат поне една поредица от три или повече високоинтензивни дни, но общият дял на тези дни спрямо всички високоинтензивни дни е само 3.8 - 4.2%. Това показва, че интензивните тренировки са кратки и изолирани, в съответствие с принципите на натоварване и възстановяване в биатлона.

Съпоставянето потвърждава, че моделът е усвоил не само количествените зависимости (разпределение по зони), но и функционалната структура на натоварването, включително редуването между работа с висока и ниска интензивност в рамките на микроцикли.

Верификацията на обучението е реализирана чрез отделен валидационен набор (приблизително 10% от програмите), върху който след всяка епоха се проследява динамиката на показателите и се предотвратява пренастройване към обучаващите данни.

В процеса на обучение е използван оптимизатор с параметри $\beta_1=0.5$ и $\beta_2=0.999$, при фиксирани скорости на обучение $\eta G=\eta D=2 \times 10^{-4}$.

Размерът на партидата е 512 дневни записа, което осигурява достатъчна стохастичност на градиентите и стабилност на сходимостта. Обучението е проведено за 80 епохи при съотношение на актуализациите $D:G = 1:1$ и с разбъркване на записите във всяка епоха, за да се поддържа равновесие между генератора и дискриминатора, както и да се избегне пренастищане на дискриминатора.

9.4.1. Проверка за логическа съгласуваност (правила върху дневни записи)

За да се удостовери, че генерираните програми следват методическата логика на тренировъчния процес, са дефинирани правила върху дневни записи, които са приложени идентично върху обучаващата и тестовата извадка.

Правило 1 (ден на състезание): изискване за липса на натоварване - Zone 1 = Zone 2 = Zone 3 = Zone 4 = Zone 5 = Strength = 0.

Правило 2 (последователни високоинтензивни дни): изискване да не се срещат повече от два поредни дни с висока интензивност, като висока интензивност е дефинирана като $\text{HighIntense} \equiv (\text{Zone 4} + \text{Zone 5}) > 6$ мин (6 мин е минимален праг за високоинтензивен компонент).

Всички проверки са извършени в минути, като правилата са приложени върху дневните записи за целия период.

Таблица 4. Резултати от проверката за логическа съгласуваност.

Правило	Показател	Обучаваща извадка (бр.)	Тестова извадка (бр.)
1. Ден на състезание - Zone 1 = Zone 2 = Zone 3 = Zone 4 = Zone 5 = Strength = 0	Брой дни	3490	401
	Случаи с натоварване	0	0
	Дял с натоварване (%)	0	0
2. Последователни високоинтензивни дни - не повече от два поредни с Zone 4 + Zone 5 > 6 мин	Брой програми	180	20
	Програми с поредица ≥ 3 дни	141	15
	Дял програми с поредица ≥ 3 дни (%)	78,33	75
	Общо високоинтензивни дни	5333	574
	Дни в поредица ≥ 3 дни	225	22
	Дял от всички високоинтензивни дни (%)	4,22	3,83

Установява се пълно съответствие с правилото за почивка. В нито един от наборите от данни не се наблюдава натоварване по която и да е зона (0% случаи с активност), което показва коректно възпроизвеждане на логиката, при която денят на контролно или основно състезание не се комбинира с тренировъчна работа.

Приблизително три четвърти от програмите в двете извадки съдържат поне една поредица от три или повече високоинтензивни дни, но общият дял на тези дни спрямо всички високоинтензивни дни е само 3,8-4,2%. Това означава, че интензивните тренировки са кратки и изолирани, като се запазва необходимото редуване между натоварване и възстановяване в рамките на микроцикъла.

Съпоставянето на резултатите показва висока степен на сходство между обучаващата извадка и генерираните програми - както по честота на поредиците, така и по относителен дял на високоинтензивните дни. Сходството потвърждава, че моделът усвоява не само количествените зависимости в данните, но и функционалната структура на натоварването.

10. ЕКСПЕРИМЕНТАЛНО ПРИЛОЖЕНИЕ: МЛАДЕЖКИ НАЦИОНАЛЕН ОТБОР ПО БИАТЛОН (СЕЗОН 2023-2024)

За проверка на ефективността на разработения генеративен модел е проведен експеримент в реални условия, при който е приложена тренировъчна програма, генерирана от модела, върху състезатели от младежкия национален отбор по биатлон през периода 2023/2024 г.

10.1. Контролни тестове и показатели

За обективна оценка са проведени входящ и изходящ контролен тест при стандартизирани условия (идентична писта, температура и състояние на настилка). Измерени са три ключови показателя:

- Максимална кислородна консумация ($\text{VO}_{2\max}$) - показател за аеробния капацитет.
- Време за 10 km бягане с ролкови ски тип Marwe 6 - интегративен индикатор за общата издръжливост, техническа ефективност и специфична скоростна издръжливост.
- Средна скорост на бягане ($\text{m}\cdot\text{s}^{-1}$) - производна от времето и разстоянието, характеризираща издръжливостта.

Използваният тип ролкови ски Marwe 6 е стандарт в контролните тестове за биатлон и ски-бягане, тъй като осигурява висока повторяемост на резултатите и механично съпротивление, близко до това на приплъзване върху снежна настилка.

10.2. Индивидуални резултати от входящ и изходящ тест

Таблица 5 представя индивидуалните резултати на осем състезатели ($n=8$), както и разликите между входящия и изходящия тест (Δ).

Таблица 5. Резултати от входящ и изходящ тест (10 km ролкови ски Marwe 6)

Биат л.	$\text{VO}_{2\max}$ вход ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$)	$\text{VO}_{2\max}$ изход ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$)	10 km вход (mm:ss)	10 km изход (mm:ss)	Δ врем е (s)	Скорос т вход ($\text{m}\cdot\text{s}^{-1}$)	Скорос т изход ($\text{m}\cdot\text{s}^{-1}$)	Δ скорос т ($\text{m}\cdot\text{s}^{-1}$)
1	60.8	63.6	27:04	24:34	-150	6.16	6.79	0.63
2	65.0	67.8	28:46	27:05	-101	5.80	6.15	0.35
3	63.6	69.0	26:38	24:14	-144	6.26	6.88	0.62
4	60.0	65.4	26:49	25:14	-95	6.22	6.61	0.39
5	61.8	66.0	27:24	25:37	-107	6.08	6.51	0.43
6	66.3	69.9	28:59	26:41	-138	5.75	6.24	0.49
7	58.1	62.1	26:52	25:20	-92	6.21	6.58	0.37
8	55.0	59.0	30:30	29:10	-80	5.47	5.71	0.24

10.3. Вариационен и статистически анализ на ефекта

Вариационният анализ показва висока хомогенност на групата ($CV < 7\%$) и физиологично консистентна динамика между трите променливи: повишаването на VO_{2max} е съпроводено с намаляване на времето и увеличаване на скоростта.

Таблица 6. Вариационен анализ на входящ и изходящ тест ($n=8$)

Показател	n	Min	Max	R	Mean	SD	Асиметрия (Sk)	Ексцес (K)	V (%)
VO_{2max} - входящ тест ($ml \cdot kg^{-1} \cdot min^{-1}$)	8	55	66.3	11.3	61.33	3.69	-0.32	-0.75	6.01
VO_{2max} - изходящ тест ($ml \cdot kg^{-1} \cdot min^{-1}$)	8	59	69.9	10.9	65.98	3.73	-0.24	-0.91	5.65
10 km - входящ тест (s)	8	1598	1830	232	1698.9	82.7	0.52	-0.94	4.87
10 km - изходящ тест (s)	8	1454	1750	296	1564.4	103.6	0.81	0.17	6.62
Скорост - входящ тест ($m \cdot s^{-1}$)	8	5.47	6.26	0.79	5.99	0.28	-0.29	-1.13	4.68
Скорост - изходящ тест ($m \cdot s^{-1}$)	8	5.71	6.88	1.17	6.43	0.40	-0.33	-1.01	6.22

За проверка на предпоставките е приложен тестът на Shapiro-Wilk ($\alpha = 0.05$) върху разликите между входящ и изходящ тест. На тази основа е избран непараметричният тест Wilcoxon за зависими извадки.

Резултатите от теста Wilcoxon показват статистически значими разлики между входящия и изходящия тест по трите показателя, с $p\text{-value} = 0.008 (< 0.05)$ за всеки от тях. Това означава, че вероятността наблюдаваните промени да се дължат на случайност е под 1%.

Размерът на ефекта е изчислен по формулата $r = Z/\sqrt{n}$, като получената стойност $r = 0.94$ ($n = 8$) показва голям ефект (по класификацията на Cohen) и силно практическо въздействие на приложената тренировъчна програма върху функционалните показатели на състезателите.

За пълнота, резултатите от проверката за нормалност (Shapiro–Wilk) и теста за значимост на разликите (Wilcoxon) са представени в Таблица 7 и Таблица 8

Таблица 7. Резултати от теста на Shapiro–Wilk за нормалност на разликите ($n=8$; $\alpha=0.05$).

Показател (разлика изходен–входящ)	p-value	Интерпретация ($\alpha=0.05$)
VO_{2max} ($ml \cdot kg^{-1} \cdot min^{-1}$)	0.047	Разпределението се различава от нормалното ($p<0.05$)
10 km време (s)	0.031	Разпределението се различава от нормалното

		($p < 0.05$)
Скорост ($\text{m} \cdot \text{s}^{-1}$)	0.039	Разпределението се различава от нормалното ($p < 0.05$)

Таблица 8. Резултати от теста на Wilcoxon за свързани извадки ($n=8$; $\alpha=0.05$).

Показател	p-value	Интерпретация ($\alpha=0.05$)
VO_2max ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$)	0.008	Статистически значима разлика – увеличение на VO_2max
10 km време (s)	0.008	Статистически значима разлика – намаление на времето
Скорост ($\text{m} \cdot \text{s}^{-1}$)	0.008	Статистически значима разлика – увеличение на скоростта

Планираното тренировъчно натоварване за сезон 2023–2024 г. обхваща периода (01.04.2023 – 31.03.2024), като за всеки ден са отчетени минутите работа в пет зони на натоварване и времето за силова подготовка, както и принадлежността към период на подготовка.

Общият обем извършена работа за разглеждания период възлиза на 34 045 минути (567.4 часа). От тях 30 695 минути (511.6 часа) са работа за издръжливост в зони 1–5, а 3 350 минути (55.8 часа) са отделени за силова подготовка. Силовата подготовка формира 9.8% от общия годишен обем и 10.9% от обема на тренировките за издръжливост.

Разпределението на времето по зони на натоварване показва ясно изразен „пирамидален“ модел: Zone 1 – 293.9 часа (57.4% от времето в зони 1–5); Zone 2 – 143.8 часа (28.1%); Zone 3 – 43.4 часа (8.5%); Zone 4 – 18.4 часа (3.6%); Zone 5 – 12.1 часа (2.4%). Сумирано, Zone 1 и Zone 2 формират 85.5% от времето в тренировките за издръжливост, докато зони 3–5 заемат 14.5%, което е типично за спортовете за издръжливост и е съвместимо с принципите на периодизацията в биатлона.

11. РЕЗУЛТАТИ И АНАЛИЗ

Резултатите от дисертационния труд са групирани в два основни направления: (1) методологични резултати - разработване и формализиране на интегрална методика и моделна архитектура за оптимизиране на тренировъчните натоварвания чрез генеративни модели; (2) емпирични резултати - статистическа и практическа верификация на модела чрез анализ на сходството между обучаващи и генерирани програми и експериментално приложение в реална тренировъчна среда.

11.1. Резултати от анализа на разпределението по зони и логиката на натоварване

Анализът на разпределението на дневните натоварвания по зони (Zone 1-5) показва закономерно доминиране на ниско- и умереноинтензивни натоварвания (Zone 1-2), което съответства на изискванията за изграждане на аеробна база и устойчивост при биатлона.

Стратификацията по VO_2max демонстрира монотонна зависимост между аеробния капацитет и обема на натоварването.

Дневната и седмичната структура на натоварването показват наличие на 62 дни без тренировъчно натоварване (16,9% от всички календарни дни), което означава 304 тренировъчни дни или средно около 5,8 тренировъчни дни седмично. Средната дневна продължителност на тренировките за сезона е 93,0 минути със стандартно отклонение 63,5 минути, минимална стойност 0 минути и максимална дневна стойност 274 минути.

Медианата на дневното натоварване е 90,0 минути, а интерквартилният интервал (25-75 перцентил) е 53,5-135,0 минути. Регистрирани са 40 дни с натоварване ≥ 180 минути (около 11% от всички дни).

Изчисленият среден седмичен обем работа е около 10,9 часа, което характеризира натоварването като сравнително високо, но умерено и съобразено с възрастта и етапа на развитие на младежките национали по биатлон.

11.2. Силова подготовка в годишния цикъл

Извършените натоварвания показват, че силовата подготовка е концентрирана в подготвителния период, където представлява около 10-11% от общото натоварване, докато в състезателния период делът ѝ намалява до около 9%, а в преходния практически отсъства. Честотата на силовите занимания е приблизително 1.8 пъти седмично в подготвителния и 1 път седмично в състезателния период. Силовата подготовка се провежда основно в дни с работа за издръжливост, което води до разграничаване на дни с по-голямо общо натоварване и дни с по-нисък общ обем и изисква прецизен контрол за ограничаване на риска от преумора.

Общият извършен обем работа за сила в изследвания период е 3 350 минути, или 55,8 часа, което съответства на 9,8% от общото тренировъчно натоварване и на 10,9% спрямо обема работа в зони на натоварване 1-5.

Силова подготовка е отчетена в 77 дни, което означава средно около 1,5 силови занимания седмично за целия сезон, със средна продължителност на една силова тренировка 43,5 минути (медиана 45 минути). Най-голям дял от този обем е реализиран в подготвителния период - 2 460 минути, или 41,0 часа (73,4% от годишния силов обем). В рамките на подготвителния период силовата подготовка формира 10,5% от общото тренировъчно натоварване и 11,7% спрямо работата в зони на натоварване 1-5. Регистрирани са 57 дни със силова подготовка, приблизително 2 силови занимания седмично, със средна продължителност 43,2 минути.

В състезателния период извършеният обем работа за сила възлиза на 875 минути (14,6 часа), или 26,1% от годишния обем по този показател. Делът на силата в общото натоварване за периода е 8,8%, а спрямо обема работа в зони на натоварване 1-5 - 9,7%. Отчетени са 19 дни със силова подготовка, което съответства на около 1 силово занимание седмично, със средна продължителност 46,1 минути. В преходния период силовата подготовка практически отсъства - реализирани са 15 минути (0,25 часа), разпределени в един ден.

За да се оцени как силовата подготовка се комбинира с тренировките за издръжливост в рамките на отделните дни, са анализирани всички дни с тренировъчно

натоварване (общо 304 дни), от които 77 дни със силова подготовка и 227 дни само с работа в зони на натоварване.

В дните със силова подготовка средният обем работа в зони на натоварване е 98,6 минути (медиана 92,0 мин; $SD \approx 48,2$ мин), а силовата подготовка е средно 43,5 минути (медиана 45,0 мин), което формира общо дневно натоварване 142,1 минути. В дните без силова подготовка, но с тренировка, средната стойност на работа в зони на натоварване е 101,8 минути (медиана 92,0 мин; $SD \approx 46,9$ мин). Обемът работа в зони на натоварване е практически еднакъв в дните със и без силова подготовка, като съществена разлика се наблюдава в общото дневно натоварване, което е приблизително с 40 минути по-високо в дните със силова подготовка.

При разглеждане само на дните със силова подготовка коефициентът на корелация на Пирсън между минутите работа за сила и минутите работа в зони на натоварване е $r = 0,39$ (умерена положителна корелация). Когато в анализа се включат всички тренировъчни дни, включително тези без силова подготовка, корелацията между дневната работа в зони на натоварване и минутите силова подготовка практически изчезва ($r \approx 0,05$) поради големия брой дни с 0 минути сила при широк диапазон на стойностите в зоните.

В седмичен диапазон разпределението на силовата подготовка по периоди се изразява приблизително в: около 1,3 часа силова подготовка и около 11,0 часа работа в зони на натоварване седмично в подготвителния период; около 0,8 часа силова подготовка и около 8,6 часа работа в зони на натоварване седмично в състезателния период; и минимален обем силова работа при около 3,9 часа седмична работа в зони на натоварване в преходния период.

Тези стойности потвърждават ясно изразена концентрация на силовата подготовка в подготвителния период, с умерено присъствие в състезателния и практически отсъствие в преходния период. В микроциклически аспект силовата подготовка се реализира предимно в дни с доминираща работа за издръжливост, без съществено намаляване на минутите в зоните на натоварване, което води до ясно разграничаване на дни с по-голямо общо натоварване и дни с по-нисък обем като част от редуването натоварване–възстановяване.

11.5. Анализ на годишния обем и периодизацията (сезон 2023-2024)

За целите на описателния анализ данните са групирани по тренировъчни периоди (подготвителен, състезателен и преходен). Изчислени са общият обем работа (в минути и часове), относителният дял на времето във всяка зона на натоварване, средни дневни стойности и стандартно отклонение, както и честота на дни без регистрирано тренировъчно натоварване.

Планираното тренировъчно натоварване за сезон 2023–2024 г. обхваща периода 01.04.2023 – 31.03.2024, като за всеки ден са отчетени минутите работа в пет зони на натоварване (Zone 1–Zone 5) и времето за силова подготовка (Strength), както и принадлежността към преходен, подготвителен или състезателен период.

Общият обем извършена работа за разглеждания период възлиза на 34 045 минути (567,4 часа). От тях 30 695 минути (511,6 часа) са работа за издръжливост в зони 1–5, а 3 350 минути (55,8 часа) са отделени за силова подготовка. Силовата подготовка представлява 9,8% от общия годишен обем и 10,9% спрямо обема на тренировките за издръжливост.

Разпределението на времето по зони на натоварване следва ясно изразен „пирамидален“ модел: Zone 1 – 293,9 часа (57,4% от времето в зони 1–5); Zone 2 – 143,8 часа (28,1%); Zone 3 – 43,4 часа (8,5%); Zone 4 – 18,4 часа (3,6%); Zone 5 – 12,1 часа (2,4%). Сумирано, Zone 1 и Zone 2 формират 85,5% от времето в тренировките за издръжливост, докато Zone 3–Zone 5 заемат 14,5%.

Дневната и седмичната структура на натоварването показват 62 дни без тренировъчно натоварване (16,9% от всички календарни дни), което означава 304 тренировъчни дни или средно около 5,8 тренировъчни дни седмично. Средната дневна продължителност на тренировките за сезона е 93,0 минути със стандартно отклонение 63,5 минути; минималната стойност е 0 минути, а максималната достига 274 минути. Медианата на дневното натоварване е 90,0 минути, а интерквartilният интервал (25–75 персентил) е 53,5–135,0 минути. Регистрирани са 40 дни с натоварване ≥ 180 минути (около 11% от всички дни). Изчисленият среден седмичен обем работа е около 10,9 часа.

Тези показатели характеризират натоварването като сравнително високо, но умерено и съобразено с възрастта и етапа на развитие на състезателите, при доминиране на ниско- и умереноинтензивната работа и контролирано присъствие на по-високите зони в съответствие с принципите на спортната периодизация.

Подготвителният период обхваща 222 дни и включва 194 тренировъчни дни, от които 28 дни без тренировка (12,6%). В този период се концентрира най-голямата част от годишното натоварване: общ обем работа 390,1 часа (68,8% от сезонния обем), обем на тренировките за издръжливост (зони 1-5) 349,1 часа и силова подготовка 41,0 часа (73,4% от годишния силов обем).

Средната дневна продължителност на натоварването в подготвителния период е 105,4 минути със стандартно отклонение 65,6 минути; средният седмичен обем е около 12,3 часа, а средната продължителност на активните тренировъчни дни е 120,6 минути.

Разпределението по зони в подготвителния период се характеризира с доминиране на зона 1 и зона 2 и поддържащ дял на високоинтензивната работа: зона 1 - 200,1 часа (57,3%); зона 2 - 101,3 часа (29,0%); зона 3 - 33,5 часа (9,6%); зона 4 - 9,0 часа (2,6%); зона 5 - 5,2 часа (1,5%). Зони 1-2 представляват 86,3% от обема на издръжливостта, а зони 3-5 - 13,7%, като зоните 4 и 5 заемат 4,1%.

Състезателният период включва 123 дни, от които 98 са тренировъчни, а 25 са без тренировка (20,3%). Реализирани са общ обем работа 165,5 часа (29,2% от годишния обем), обем на тренировките за издръжливост 150,9 часа и силова подготовка 14,6 часа (26,2% от годишния силов обем).

Средната дневна продължителност в състезателния период е 80,7 минути със стандартно отклонение 55,0 минути; средният седмичен обем е около 9,4 часа, а средната продължителност на активните тренировъчни дни е 101,3 минути. В сравнение с подготвителния период се наблюдава намаляване на общия обем и увеличаване на дела на почивните дни.

Разпределението по зони в състезателния период е: зона 1 - 87,9 часа (58,2%); зона 2 - 37,9 часа (25,1%); зона 3 - 9,3 часа (6,2%); зона 4 - 8,9 часа (5,9%); зона 5 - 7,0 часа (4,6%). Делът на високите зони 4 и 5 нараства до 10,5% (срещу 4,1% в подготвителния период) при по-нисък общ обем, което е в съответствие с логиката на подготовка за състезания.

Преходният период включва 21 дни, от които 12 са тренировъчни, а 9 дни са без натоварване (42,9%). Общият обем натоварване е 11,8 часа (2,1% от годишния обем),

като работата за издръжливост е 11,6 часа, а силовата подготовка е 0,2 часа (12 минути). Средната дневна стойност е 33,7 минути със стандартно отклонение 35,0 минути, средният седмичен обем е около 3,9 часа, а средната продължителност на активните тренировъчни дни е 59,0 минути.

Разпределението по зони в преходния период е: зона 1 - 5,9 часа (50,9%); зона 2 - 4,6 часа (40,0%); зона 3 - 0,6 часа (5,3%); зона 4 - 0,4 часа (3,8%); зона 5 - 0 часа, което потвърждава характера на периода като „активно възстановяване“ с доминация на най-ниските зони.

В обобщение, описателният анализ показва ясно структурирана годишна периодизация: подготвителният период концентрира около две трети от годишния обем при преобладаващо ниска до умерена интензивност; състезателният период се характеризира с по-нисък общ обем и относително нарастване на високите зони; преходният период е фаза на възстановяване с минимален обем, доминация на зони 1-2 и практически отсъствие на работа в зона 5 и силова подготовка.

Таблица 9. Обобщение на годишния обем работа по тренировъчни периоди (сезон 2023-2024).

Период	Календарни дни	Тренировъчни дни	Дни без тренировка (%)	Обем общо (часа)	Издръжливост / Сила (часа)
Подготвителен	222	194	28 (12,6%)	390,1	349,1 / 41,0
Състезателен	123	98	25 (20,3%)	165,5	150,9 / 14,6
Преходен	21	12	9 (42,9%)	11,8	11,6 / 0,2

Резултатите в таблица 9 показва концентрация на обема работа в подготвителния период и закономерно редуциране на натоварването в състезателния и преходния период, при едновременно нарастване на относителния дял на високите зони в състезателната фаза.

11.6. KL-дивергенция и сходство на разпределенията по зони

За оценка на сходството между реалните и генерираните разпределения по зони на натоварване е изчислена KL-дивергенция (Kullback-Leibler) по квантилни групи на $\text{VO}_{2\max}$. Във всяка подгрупа сумарните обеми по зони (Zone 1 - Zone 5 и Сила, в минути) са нормализирани като относителни дялове от общия дневен обем и са сравнени между реалните (P) и генерираните (Q) данни.

Получените стойности на $\text{KL}(P||Q)$ са от порядъка на 0,0001-0,0004 във всички квантилни групи, което е индикатор за изключително близка форма на разпределенията по зони между реалните и генерираните данни. Сходството е равномерно в Q1-Q4, без квантил с отчетливо по-висока KL, което показва стабилно условно поведение на модела спрямо $\text{VO}_{2\max}$.

Таблица 10. KL-дивергенция между реални и генерирани разпределения по квантили на VO₂max.

VO ₂ -квантил	Q1	Q2	Q3	Q4
KL(P Q)	0,0003	0	0	0,0001

Допълнителна интерпретация на сходството е представена чрез относителните дялове по зони (P - реално, Q - генерирано), като по-долу са включени редове, представени в дисертационния труд като част от всички стойности.

Таблица 11. Относителни дялове по зони (P - реално, Q - генерирано) (представени са само част от всички редове).

VO ₂ -кв.	Зона	P(реално)	Q(генерирано)
Q1	Zone 1	0.5482	0.5513
Q2	Zone 1	0.5473	0.5459
Q3	Zone 1	0.5484	0.539
Q4	Zone 1	0.5474	0.5523
Q1	Zone 2	0.244	0.2473
Q2	Zone 2	0.2432	0.2423
Q3	Zone 2	0.2449	0.2458
Q4	Zone 2	0.2447	0.2427
Q1	Zone 3	0.0752	0.0753
Q2	Zone 3	0.0765	0.0769
Q3	Zone 3	0.0758	0.0776
Q4	Zone 3	0.0775	0.0774
Q1	Zone 4	0.0261	0.0235
Q2	Zone 4	0.026	0.0269
Q3	Zone 4	0.0252	0.028
Q4	Zone 4	0.0245	0.0236
Q1	Zone 5	0.0184	0.0165
Q2	Zone 5	0.0178	0.0192

От обобщените резултати следва, че моделът правилно разпределя общия дневен обем по зони на натоварване в рамките на всяко ниво на тренираност. Най-малки различия се наблюдават в Zone 3 и Zone 5, докато в Zone 1-2 отклоненията са минимални (в рамките на 0,002-0,010), което е очаквано предвид доминиращия им дял.

11.7. Логическа съгласуваност

Логическата верификация допълва количествените метрики чрез формализирани правила за контрол на вътрешната съгласуваност и методическата допустимост на генерираните натоварвания. Проверени са ограничения върху дневните записи, свързани с липса на тренировъчно натоварване в ден на състезание и ограничаване на последователни високоинтензивни дни при дефиниция $\text{HighIntense} \equiv (\text{Zone 4} + \text{Zone 5}) > 6 \text{ мин.}$

Резултатите от проверката показват пълно съответствие с правилото за почивка (0% случаи с активност в ден на състезание) и висока степен на структурна устойчивост по отношение на редуването между натоварване и възстановяване в рамките на микроцикъла. Макар значителна част от програмите да съдържат поне една поредица от

три или повече високоинтензивни дни, относителният дял на тези дни спрямо всички високоинтензивни дни остава малък (3,8-4,2%), което характеризира интензивните натоварвания като кратки и изолирани.

11.8. Обобщение на статистическата верификация

Комбинацията от вариационен и квантилен анализ, мерки за сходство на разпределенията (KL-дивергенция), метрики за грешка (САГ), тестове за нормалност (Shapiro-Wilk) и значимост (Wilcoxon) и оценка на размера на ефекта (r) осигурява пълна количествена и структурна оценка на резултатите и е методологично адекватна за данните от изследването.

В Таблица 12 и Таблица 13 са представени ключови количествени показатели за точността на възпроизвеждане на обучаващите данни от генеративния модел чрез средна абсолютна грешка (САГ), изчислена в минути по енергийни режими и по квантилни групи на VO_{2max} (Q1–Q4).

Таблица 12. САГ по зони на натоварване (в минути) и квантилни групи по VO_{2max} .

VO_{2max} -квантил	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Сила	Средна САГ
Q1 (ниска тренираност)	1.09	0.64	0.12	0.17	0.12	0.02	0.36
Q2 (умерена тренираност)	1.46	0.67	0.15	0.01	0.08	0.24	0.43
Q3 (висока тренираност)	0.51	0.72	0.37	0.35	0.20	0.46	0.44
Q4 (много висока тренираност)	0.49	0.23	0.02	0.10	0.11	0.10	0.18

Таблица 13. Обобщени стойности на САГ по зони на натоварване (в минути).

Зона на натоварване	Средна САГ (мин.)
Zone 1	0.89
Zone 2	0.57
Zone 3	0.16
Zone 4	0.16
Zone 5	0.13
Сила	0.20
Средна обща САГ	0.35

Резултатите показват висока точност на възпроизвеждане на количествените зависимости между тренировъчните режими и нивото на тренираност. Средната обща САГ е 0.35 минути, което съответства на под 1% отклонение спрямо средните стойности на натоварванията в обучаващата съвкупност. Същевременно, ниските стойности на САГ при високоинтензивните зони и силовата подготовка потвърждават, че моделът възпроизвежда и по-редките компоненти на натоварването, без да нарушава микроцикличната логика.

11.4. Практическа приложимост

Експерименталното приложение на генерирана тренировъчна програма за сезон 2023-2024 г. доказва наличието на реален тренировъчен ефект, изразен в статистически значимо подобрене на $\dot{V}O_{2\max}$, съкращаване на времето в 10 km тест и увеличение на средната скорост. Това утвърждава оперативната ефективност на модела и потенциала му за внедряване като инструмент за анализ, планиране и прогнозиране на натоварването.

12. ЗАКЛЮЧЕНИЕ

Интегрирането на генеративни модели на изкуствения интелект със съвременната спортна теория представлява методологично последователен и емпирично валидиран подход за оптимизиране на тренировъчния процес в цикличните спортове.

- Тренировъчният процес може надеждно да бъде формализиран като многокритериална оптимизационна задача, при която целевата функция отразява адаптационния ефект, а управляващите параметри са обемът, интензивността и времето за възстановяване.
- Предложената интегрална методика от взаимосвързани етапи (проектиране, агрегиране, обработка, управление, оценка) функционира като „затворена“ система за управление, в която данните постоянно усъвършенстват модела и осигуряват възпроизводимост и проследимост.
- Разработеният генеративен модел възпроизвежда количествените и структурни зависимости на тренировъчния процес и генерира програми, съпоставими по структура и динамика с експертно създадените.
- Експерименталното приложение при младежкия национален отбор по биатлон показва статистически значимо подобрене на функционалните показатели и много голям практически ефект, което потвърждава приложимостта на модела в реална спортна среда.
- Оптимизацията следва да бъде многокритериална с ясно дефиниран вектор от цели; целесъобразно е класическите методи за оптимизация да се използват за локално оптимизиране, а генеративните модели - за глобален синтез и селекция на тренировъчни сценарии.
- Препоръчва се периодична актуализация на параметрите в края на микро- и мезоцикли на базата на новопостъпили данни (показатели за умора, възстановяване и др.), както и разширяване на входните данни към мултимодални източници (физиологични, биомеханични и психологически показатели).

Разработената в дисертационния труд концепция показва, че генеративният подход позволява тренировъчните програми да бъдат представени като многомерни структури, чиято логика може да се възпроизвежда чрез обучение върху експертно разработени годишни планове. В този модел тренировъчното натоварване се разглежда като управляем обект, а параметрите на управлението се описват чрез обем, интензивност, времева структура и ограничения, произтичащи от периодизацията.

Емпиричната верификация потвърждава, че качеството на генерираните програми може да бъде оценявано надеждно чрез комбинация от количествени и логически критерии.

Ниските стойности на средната абсолютна грешка (САГ) по енергийни режими, изключително малките стойности на KL-дивергенцията между реални и генерирани разпределения по зони и резултатите от логическата проверка на микроцикличната структура показват, че моделът възпроизвежда както количествените зависимости (разпределение по зони), така и методическата последователност на натоварване–възстановяване.

В контекста на младежкия национален отбор по биатлон (сезон 2023–2024) практическата приложимост на предложения подход е подкрепена от резултатите от входящ и изходящ контролен тест, както и от статистическия анализ на ефекта. Подобриенето на $\dot{V}O_{2\max}$, редуцията на времето в 10 km тест и увеличението на средната скорост отразяват наличие на реален тренировъчен ефект и показват, че генерираната програма не е само статистически правдоподобна, но и функционално адекватна в реална спортна среда.

От методологична гледна точка ключовото преимущество на интегралната методика е, че функционира като „затворена“ система за управление: данните от реалната подготовка захранват генеративния модел, резултатите се оценяват по формализирани критерии, а параметрите се актуализират периодично в края на микро- и мезоцикли. Тази логика позволява адаптивност на решенията при промяна в функционалното състояние, календарния контекст и целите на подготовката.

Практическото внедряване на подхода предполага спазване на два задължителни принципа. Първо, генеративният модел следва да бъде ограничаван чрез методически правила, които да предотвратяват физиологично недопустими комбинации на натоварването. Второ, оценката на резултатите следва да бъде многокритериална и да включва както количествени метрики (САГ, KL-дивергенция), така и логическа съгласуваност на микроцикличната структура и контрол върху съдържанието на натоварването в специфични дни (напр. преди контролно състезание).

Ограничение на изследването произтича от факта, че генеративното обучение възпроизвежда закономерности, налични в обучаващия масив. Следователно, точността и устойчивостта на модела са пряко зависими от представителността, качеството и консистентността на тренировъчните програми, използвани за обучение, както и от обхвата на входните променливи.

Перспективите за развитие на методиката включват: разширяване на входните данни към мултимодални източници (физиологични, биохимични и функционални показатели), интеграция на показатели за умора и възстановяване, както и разширяване на логическите правила за контрол на методическата допустимост в условията на по-голямо разнообразие от тренировъчни средства и състезателни сценарии.

В обобщение, дисертационният труд показва, че Generative AI може да бъде използван не като автономен заместител на треньора, а като инструмент за подпомагане на планирането и управлението на натоварването чрез систематично генериране на алтернативи, оценка на тяхното качество и формализирано приваждане към методическите принципи на спортната тренировка.

13. НАУЧНИ И ПРИЛОЖНИ ПРИНОСИ

13.1. Научни приноси

1. Разработена е интегрална методика за оптимизиране на тренировъчния процес чрез генеративни модели на изкуствения интелект, в която е представена последователна структура от етапи – проектиране на модела, събиране и обработка на данни, управление на генеративния алгоритъм и оценка на ефективността – която дефинира системен подход за прилагане на AI-технологии в спортната подготовка.

2. Обосновано е използването на GAN архитектура като средство за генериране на тренировъчни програми.

13.2. Приложни приноси

3. Демонстрирана е възможността чрез генеративен модел на изкуствен интелект да бъдат възпроизведени тренировъчни програми с минимални отклонения спрямо реалните, използвани за обучение на модела.

4. Приложени са система от формализирани логически правила за контрол на вътрешната съгласуваност и методическата допустимост на генерираните от модела тренировъчни натоварвания.

5. Разработена е тренировъчна програма за сезон 2023–2024 г., базирана изцяло на генеративен модел, и приложена за целите на младежкия национален отбор по биатлон.

**NATIONAL SPORTS ACADEMY “VASIL LEVSKI”
DEPARTMENT OF “SPORT THEORY”**

ABSTRACT

of a dissertation submitted for the award of the scientific degree

“DOCTOR OF SCIENCE (DSc)”

in Professional Field 7.6 “Sport”

**AN INTEGRAL METHODOLOGY FOR OPTIMISATION OF
TRAINING LOAD THROUGH
GENERATIVE
ARTIFICIAL INTELLIGENCE MODELS**

Prof. Mihail Ivaylov Konchev, PhD

Sofia, 2026

DETAILS OF THE DISSERTATION

The dissertation comprises 281 pages. It contains 32 figures and 16 tables, as well as 4 appendices.

The public defence of the dissertation will take place on 17 March 2026 in Hall A-3 of the National Sports Academy “Vasil Levski”.

LIST OF ABBREVIATIONS AND SYMBOLS USED

- AI – artificial intelligence (Artificial Intelligence).
- Generative AI – generative artificial intelligence models for the synthesis of new data/structures.
- GAN – Generative Adversarial Network: generator and discriminator in a minimax formulation.
- VAE – Variational Autoencoder: probabilistic latent representation and KL regularisation.
- KL divergence – an information measure of the difference between two probability distributions (Kullback–Leibler).
- MAE – mean absolute error: the mean deviation between observed and generated values across training-load regimes/zones.
- VO_2max – maximal oxygen uptake ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) as an indicator of aerobic capacity.
- Zone 1–Zone 5 – training load zones (time in hours/minutes) describing the distribution of training volume across energy regimes.
- Control_days / Main_days – remaining days until a control / main competition as contextual (conditional) variables.

CONTENTS

1. INTRODUCTION
2. RELEVANCE AND SIGNIFICANCE OF THE STUDY
3. METHODOLOGICAL FOUNDATIONS OF THE STUDY
4. METHODOLOGICAL BASIS AND RESEARCH DESIGN
5. THEORETICAL FRAMEWORK: TRAINING LOAD AS AN OPTIMISATION PROBLEM
6. OPTIMISING TRAINING LOAD THROUGH GENERATIVE ARTIFICIAL INTELLIGENCE MODELS
7. AN INTEGRAL METHODOLOGY FOR OPTIMISING TRAINING LOAD THROUGH GENERATIVE ARTIFICIAL INTELLIGENCE MODELS
8. EMPIRICAL BASIS, DATA AND PARAMETERISATION OF THE GENERATIVE MODEL
9. GENERATIVE MODEL DESIGN AND VERIFICATION
10. EXPERIMENTAL APPLICATION: YOUTH NATIONAL BIATHLON TEAM (SEASON 2023–2024)
11. RESULTS AND ANALYSIS
12. CONCLUSION
13. SCIENTIFIC AND APPLIED CONTRIBUTIONS

1. INTRODUCTION

Optimising training loads is among the most complex methodological tasks in the theory and methodology of sports training. The training process is characterised by a high degree of dynamism, multi-component structure and the presence of factor and domain constraints arising from the athlete's functional capacities, the stage of preparation and the logic of periodisation.

In practice, load is viewed as a complex of parameters that must be quantitatively specified and managed so as to achieve a maximum functional effect while maintaining physiological stability and controlling the risk of overreaching and injury. This implies formulating an optimality criterion (objective function), selecting control parameters and applying algorithms to search for an optimum within the admissible solution space.

The development of generative artificial intelligence models makes it possible to treat training programmes as realisations of a multidimensional probability distribution derived from expert-designed and empirically validated plans. In this way, an approach is established for global exploration and synthesis of training loads, followed by control and statistical verification.

2. RELEVANCE AND SIGNIFICANCE OF THE STUDY

The relevance of the dissertation arises from the objective need for the training process to be managed as a dynamic, multi-criteria system, in which decisions on load must account for constraints, adaptation and the periodisation of training. Under conditions of high sporting mastery, even small changes in volume and intensity may have a substantial impact on the body's functional and structural responses.

Our time is marked by rapid development and widespread adoption of artificial intelligence, including generative models, which are establishing themselves as a leading field globally and in Bulgaria. In sports science, this development creates the prerequisites for a transition from empirical and partial optimisation to systematic, data-driven and reproducible management of training loads.

The significance of the study is determined by the proposed integral methodology, which integrates: (1) classical propositions of sports methodology (magnitude, character, orientation; dose–response; load and recovery), (2) mathematical modelling and optimisation algorithms, and (3) generative artificial intelligence models subject to statistical and logical verification.

In the dissertation, training preparation is examined within contemporary systems of sports preparation, where high performance is formed as the result of the cumulative influence of genotypic, phenotypic, material and technical, managerial and scientific–technological factors. In this context, training load is the primary control factor through which a purposeful impact on functional adaptations and sporting form is realised.

The topicality of the theme is further strengthened by the increasing availability of data and by the capacity of Generative AI to synthesise structures that preserve the regularities of empirical programmes, while also enabling systematic exploration of the admissible space of control parameters. This creates prerequisites for moving from local and partial solutions to a reproducible integral methodology, in which optimisation decisions are supported by statistical assessment and logical verification.

For Bulgarian sports practice, the significance of the integral methodology proposed in the dissertation is expressed in its applicability with limited samples and in the context of youth sport, where the need for precise load management is combined with a higher risk of methodological errors. The verification module with formalised rules restricts the generation of physiologically inadmissible combinations and maintains compliance with the principles of sports methodology.

3. METHODOLOGICAL FOUNDATIONS OF THE STUDY

The aim of the dissertation is to propose an integral methodology for optimising the training process through generative artificial intelligence models, and to validate the proposed methodology using data from the youth national biathlon team.

3.1. Scientific thesis

The dissertation defends the thesis that integrating generative artificial intelligence models into the planning and management of the training process enables systematic, adaptive and multi-criteria optimisation of loads in cyclical sports.

3.2. Research tasks

1. To analyse the theoretical foundations of the training process as an object of optimisation and to systematise key concepts for planning, structuring and managing training loads.
2. To conduct a critical analysis of existing methods and algorithms for optimisation applicable to training loads, including methods of systems–structural analysis, mathematical modelling and adaptive algorithms for multi-criteria optimisation.
3. To analyse the applicability of generative artificial intelligence models for optimising training loads and to assess their advantages and limitations.
4. To propose a conceptual model for integrating generative models into the training process, including a methodological framework for data collection, structuring and processing, and stages of adaptive optimisation.
5. To develop an integral model for optimising training load through generative artificial intelligence models.
6. To apply the empirically developed model using biathlon data and to generate a training programme under real-world conditions.
7. To assess the applicability of the proposed integral model by means of mathematical and statistical methods.

3.3. Object and subject of the study

The object of the study is training load in cyclical sports as a process of planning, adaptation and optimisation.

The subject of the study is the development and validation of an integral methodology for optimisation of the training process through generative artificial intelligence models, including algorithms for adaptive load management, multi-criteria optimisation and verification via mathematical and statistical methods.

4. METHODOLOGICAL BASIS AND RESEARCH DESIGN

The methodological basis of the study combines a systems–structural approach, mathematical modelling and optimisation, as well as generative artificial intelligence models intended for the synthesis and selection of training scenarios.

4.1. Research methods

The methods applied are organised into three main groups:

- Theoretical and methodological analysis: a systematic review and critical analysis of theories and algorithms for optimising training loads and generative models (VAE, GAN, etc.).
- Expert assessment: expert involvement in generating and structuring training programmes for the training and test samples, as well as in defining the input parameters (VO₂max_input, Control_days, Main_days) and methodological constraints.
- Mathematical and statistical methods: variance analysis; quantile analysis and stratification by VO₂max (Q1–Q4); mean absolute error (MAE) and KL divergence for assessing structural similarity; logical rules to control internal consistency; the Shapiro–Wilk test for normality; the Wilcoxon test for paired samples; effect size (r); and analysis of relationships via correlations and conditional probabilities.

4.2. Research design

The experimental work was implemented through a sequence of research stages: a theoretical stage (literature review and methodological basis), development of an integral methodology, expert generation of training programmes for model training, statistical evaluation of the results, and synthesis/discussion.

The design of the study is oriented towards constructing a control cycle in which: (1) empirical data from expert-planned programmes are structured and standardised; (2) the generative model is trained and validated on a separate dataset; (3) generated programmes undergo quantitative evaluation (MAE, KL divergence) and logical/physiological checking; (4) the results are used to tune model parameters and to formulate applied recommendations.

The empirical base comprises 200 individual training programmes of varying duration, represented as 44,694 daily observations. Each daily record contains input indicators (VO₂max_input and contextual variables to competition), as well as output indicators—time by training load zones (Zone 1–Zone 5) and a “Strength” module, expressed in hours/minutes. The structure enables quantile analysis and stratification by level of preparedness (Q1–Q4).

The statistical framework includes descriptive analysis (central tendency, variability, skewness and kurtosis), tests of normality (Shapiro–Wilk), non-parametric tests for paired samples (Wilcoxon), as well as assessment of practical effect (effect size r and test power). This framework was used both to evaluate similarity between training and generated programmes and to verify the effect of the experimental application.

Data quality control was also applied (validation of types, ranges, missing values and consistency between variables), as well as standardisation of input and output parameters to ensure stable training and to limit the risk of “mode collapse” in the generative architecture.

5. THEORETICAL FRAMEWORK: TRAINING LOAD AS AN OPTIMISATION PROBLEM

5.1. Conceptual apparatus and structuring of load

In the dissertation, training load is considered as a multi-component system characterised by magnitude, character and orientation. Magnitude is described by volume and intensity; character—by coordination complexity and degree of similarity to competitive activity; orientation—by the objective function of the impact on vegetative, motor and psychomotor functions. The hierarchical principle and deductive approach allow load to be considered at different levels, reaching quantitatively measurable parameters required for formalisation and precise dosing.

A central role is played by the dose–response relationship and the range of impact within which load acts as a formative stimulus, inducing positive functional and structural changes. Dosing criteria define different categories of load (maximal, medium, small) and the need for dynamic balance among them in the training process.

5.2. Mathematical formulation of the optimisation problem

Training load is formalised as a controllable object described by a vector of control parameters $x = (x_1, x_2, \dots, x_n)$, which may include volume, intensity, frequency, recovery intervals, etc. The optimality criterion (objective function) is considered as $Q(x)$, and the optimisation problem—as the search for an extremum in the admissible space Γx in the presence of factor and functional constraints.

In general form, the problem is stated as: finding $x^* \in \Gamma x$ such that $Q(x^*) = Q_{\max}$ (or Q_{\min}), subject to equality and inequality constraints. When the objective function and constraints are linear, the problem falls under linear programming; with non-linear relationships—under non-linear programming. In the context of training loads, multi-criteria optimisation and the use of a vector criterion are often required, due to the simultaneous management of competing goals (performance, recovery, control of injury risk, etc.).

5.3. Specific features of objective functions in training practice

Objective functions may be unimodal or multimodal, convex or concave, as well as “valley-shaped” and plateau-like. The multi-extremal nature of $Q(x)$ reflects the presence of multiple local optima, driven by non-linear physiological relationships and interactions among the control parameters. Valley behaviour occurs with strongly differing scales and parameter sensitivities, which hinders convergence and the reliability of classical local methods.

5.4. Algorithms for optimising complex objective functions

The dissertation systematises key requirements for optimisation algorithms: convergence, rapid convergence, low working-memory requirements, easy preparation of conditions, reliable stopping criteria and compliance with constraints. In practice, no universal algorithm exists that applies to all classes of problems; therefore, selection depends on the shape of the objective function, the number of parameters and the constraints.

5.4.1. The Gelfand–Tsetlin method

The method implements a valley search for an extremum with two or more control parameters by localising the extremum along the gradient direction from two initial points and subsequently moving along the valley line. In the context of cyclical sports, the method is interpreted through the management of parameters such as volume, intensity and functional indicators (HR, LA, etc.), while observing admissible physiological limits.

5.4.2. The Hooke and Jeeves method (pattern search)

The method is a derivative-free direct-search approach involving exploratory moves and pattern moves. The main hypothesis is local invariance of the search direction, which enables efficient movement along straight valleys. In the training context, it is interpreted as successive exploration of parameters with $\pm\Delta x_i$ and selection of a configuration that improves the objective function (e.g., fatigue indicators).

5.4.3. The Rosenbrock method (rotating coordinate system)

The method uses rotation of the coordinate system so that one of the new axes coincides with the direction of the fastest improvement. This facilitates localisation of the extremum in the rotated space. In the sports interpretation, the method supports the search for optimal combinations of interrelated parameters (e.g., frequency and strength) when the level curves of the objective function are inclined relative to the original coordinate axes.

5.4.4. Heuristic and stochastic methods for global search

For multi-extremal functions, heuristic and stochastic approaches are considered, including the Price method (cluster analysis in a population of random points) and the Wang and Luus method (pseudo-one-dimensional random search). These methods aim to overcome local optima through global exploration of the parameter space, while imposing high computational demands as dimensionality increases.

5.4.5. Gradient approaches for overcoming local optima

In the dissertation, the “heavy ball” method is examined, which introduces an inertial component (mass β) into the recurrent formula of gradient motion in order to jump over weakly expressed local minima. Such a mechanism can support optimisation on multimodal surfaces but does not guarantee reaching a global optimum from a single starting point.

5.4.6. The tunnelling method for global optimisation

The tunnelling method alternates a minimisation phase (local optimisation of $Q(x)$) and a tunnelling phase in which an auxiliary function $T(x)$ is defined to steer the search towards new regions of the admissible space, so as to avoid local optima. In the illustrative formulations, the method demonstrates a logic of successive improvement while satisfying constraints.

The tunnelling method is considered as a heuristic algorithm for global optimisation of multi-extremal objective functions, where standard local procedures often become trapped in local optima. The method provides a mechanism for controlled departure from the region of the current local minimum by constructing a tunnelling function that suppresses already-found solutions and directs the search towards new regions of the admissible space.

The algorithm includes the following main steps:

1. Selection of an initial point x_0 in the admissible solution space.

2. Minimisation phase: finding a local minimum x^* of the objective function $Q(x)$ using a local-search method (gradient-based or derivative-free), subject to the constraints.
3. Construction of a tunnelling function $T(x)$, which depends on $Q(x)$ and the current best solution x^* (respectively the value $Q^* = Q(x^*)$).
4. Tunnelling phase: finding a point $x(T)$ at which $T(x(T)) = 0$, or $T(x(T)) \approx 0$ in a numerical solution, with the aim of reaching a region with a better value of $Q(x)$ than Q^* .
5. Update: a new starting point $x_0 = x(T)$ is adopted and the algorithm is repeated from the minimisation phase.
6. Stopping criterion: alternating the two phases continues until a specified criterion is met (e.g., no improvement or reaching a pre-specified accuracy).

In the dissertation, an illustrative variant of a tunnelling function is presented, where $T(x) = Q(x) - Q^*$ for $Q(x) > Q^*$ and $T(x) = 0$ otherwise. This formulation requires, in the tunnelling phase, finding a point for which $Q(x(T)) < Q^*$, which guarantees transition to a better solution and subsequent local refinement.

The theoretical interpretation of the method shows that the tunnelling phase acts as a mechanism for global search, whereas the minimisation phase provides refinement in a local sense. This logic is compatible with optimisation problems involving constraints and competing criteria, which are characteristic of modelling training loads.

The theoretical analysis substantiates the need for classical optimisation methods to be complemented by generative artificial intelligence models that provide global synthesis and selection of training scenarios under a large number of degrees of freedom.

In the dissertation, training load is formulated as a problem of optimising control parameters that describe the magnitude, character and orientation of the impact over time. The control vector includes components for volume (time), intensity (zonal structure), frequency and alternation of load and recovery, while accounting for positioning relative to control and main competitions.

The general mathematical formulation is presented via an objective function (or a vector of objective functions) and a system of constraints reflecting physiological admissibility and methodological principles. In scalar form, the problem may be expressed as: $\min_u f(u)$ subject to $g_j(u) \leq 0$, $h_k(u) = 0$, $u \in U$, where U is the admissible space of control parameters, and g_j and h_k specify constraints related to limiting volumes, intensities, alternation of regimes and calendar constraints.

In multi-criteria optimisation, the goal is specified as a vector $F(u) = (f_1(u), f_2(u), \dots, f_m(u))$, for example: maximising the functional effect, minimising the risk of overreaching/injuries, preserving technique and coordination complexity, and minimising deviation from methodological norms. The solution is sought within the class of Pareto-efficient solutions, from which a specific programme is selected according to priorities and constraints (Pareto front).

The theoretical analysis shows that, in real training programmes, objective functions are complex, non-linear and often non-differentiable, with multiple local optima. This substantiates the use of classical derivative-free and hybrid methods (Hooke–Jeeves, Rosenbrock, Price, tunnelling method, etc.) for local refinement, as well as the need for global exploration of the solution space via generative models that synthesise candidates with preserved structural logic.

An important conclusion from the theoretical section is that optimisation in sport cannot be reduced to a single indicator or to a specific moment in time. Therefore, emphasis is placed on

microcycle and mesocycle structure, on the sequence of load and recovery, and on periodisation, which provides a context for interpreting each daily load as an element of a dynamic system.

6. OPTIMISING TRAINING LOAD THROUGH GENERATIVE ARTIFICIAL INTELLIGENCE MODELS

6.1. Theoretical foundations and types of generative models

Generative models are considered as algorithms for modelling an unknown probability distribution $p(x)$ of the observed data and for subsequently generating new realistic realisations x_{new} . In the dissertation, a comparative analysis is conducted of the main architectures: variational autoencoders (VAE), generative adversarial networks (GAN), diffusion models and normalising flows, examining their training principles and applicability to time series and structured data in sport.

VAEs integrate an encoder–decoder structure with a probabilistic interpretation via the ELBO, whereas GANs implement a competitive mechanism between a generator and a discriminator, leading to synthesis of highly realistic data. Diffusion models use a stepwise process of adding and removing noise, while normalising flows use invertible transformations that enable direct computation of densities.

Variational autoencoders compress the input data into a latent space z and subsequently reconstruct them through decoding, thereby modelling complex dependencies at reduced dimensionality. The probabilistic formulation makes it possible to work with distributions of latent variables and to generate new realisations by sampling and decoding.

In generative adversarial networks, training is based on an adversarial dynamic in which the generator seeks to synthesise realisations indistinguishable from the real data, while the discriminator evaluates whether a given realisation is real or generated. The dissertation emphasises that this mechanism can yield high “realism” of synthesis but requires careful tuning and control of training stability.

Flow-based models (normalising flows) construct complex probability distributions through a sequence of invertible transformations of a base distribution. Invertibility allows direct computation of densities and evaluation of the probability of each generated realisation, which is particularly useful in tasks requiring quantitative control over the “plausibility” of proposed solutions.

Diffusion models are described as a process of gradually introducing noise into the data and subsequently recovering the original structures via a reverse (denoising) process implemented in steps. This principle supports training stability and is applicable not only to images but also to structured data and time series when the goal is to reproduce complex empirical regularities.

The dissertation also examines the conditional formulation of generative models, in which they are driven by contextual variables. In the training context, this makes it possible to model a conditional distribution $p(x|y)$, where y includes physiological and calendar inputs (e.g., VO_2max and remaining days to a control or main competition), and x represents the structure of daily load by zones and strength training.

In this way, the training programme can be interpreted as a structured time series, in which each daily record is described by a vector of loads across intensity zones and a strength component. Generative models reproduce periodisation regularities, microcycle repeatability and constraints on admissible combinations, thereby providing a basis for integral modelling and optimisation.

6.2. Advantages and limitations of the generative approach in the training context

A key advantage of generative models is their ability to generate multiple alternative training scenarios that satisfy factor and domain constraints, thereby supporting coaching decisions under a wide range of degrees of freedom. The generative approach enables virtual “experimentation” in the parameter space, including exploration of rare or extreme combinations without immediate risk to athletes.

Limitations are related to the quality and scope of data, the computational complexity of training, risks such as mode collapse in GANs, and difficulties in interpreting models as a “black box”. This necessitates a stringent verification mechanism and expert assessment, as well as integration of logical rules and methodological filters.

In the dissertation, the generative stage is treated as a component for global search that can be combined with subsequent local refinement and control. A sequence of steps typical of “Global–Local Synergy” is summarised, progressing from training a generative model and large-scale sampling of the probability space to local fine-tuning along the volume–intensity dimension and periodic verification of solutions within a microcycle.

As a topical direction, the “Tunnel Generative + Adaptive Search” approach is discussed, combining GAN training, “tunnel” progressive narrowing of the set of candidate solutions (tunnel layers) and adaptive search through continuous model updating with signals from real-world application. This logic emphasises that generation should not be a one-off act, but a process with feedback and dynamic checking of scenario applicability under changing conditions.

In the training context, this means that the generative model should be integrated with a system for logical and physiological verification that controls internal programme consistency, compatibility between zones and strength work, and admissible sequences of consecutive high-intensity days. In this way, the risk of generating methodologically inadmissible loads is minimised while preserving statistical proximity to the training data.

Table 1. Comparison of the main classes of generative models examined in the dissertation.

Model class	Key mechanism	Strengths in the training context	Limitations and risks
VAE	Encoder–decoder with latent space z and a probabilistic formulation (ELBO).	Robust training on structured data; possibility for conditional generation and control via latent factors; convenient sampling of admissible scenarios.	Possible “averaging” of rare regimes; latent-space parameterisation and reconstruction control are required.
GAN	Adversarial generator–discriminator	High realism; effective for complex distributions and for generating	Risk of instability and mode collapse; sensitivity to a large

	process for synthesising realisations.	diverse candidate solutions for load planning.	number of parameters; need for regularisation and strict monitoring.
Diffusion models	Stepwise process of adding noise and reverse denoising to recover structures.	Stable training principle; reproduction of complex regularities; applicability to time series and structured data.	High computational complexity; more steps during generation; dependence on process tuning and the stopping criterion.
Normalising flows	A sequence of invertible transformations of a base distribution with direct density estimation.	Exact evaluation of the probability of generated realisations; control over plausibility and comparison of distributions.	Constraints due to the invertibility requirement; architectural trade-offs under strongly non-linear dependencies and high dimensionality.

The analysis shows that the choice of architecture should be aligned with the model's purpose (structural similarity of distributions, diversity of candidate solutions, reproducibility of microcycle regularities) and with the available volume and quality of data. Regardless of the choice, application in sports practice requires integration with assessment and verification criteria that guarantee methodological admissibility and physiological compatibility of the generated loads.

6.3. Compatibility of multi-criteria algorithms with generative models

In the dissertation, optimisation is considered as a multi-criteria task in which competing goals are balanced (sport performance, recovery, control of injury risk, rational use of time). Generative models are used to create a rich initial population of candidate solutions, which are subsequently evaluated and refined by multi-criteria algorithms and Pareto-oriented analysis.

The proposed “generation–evaluation–optimisation” logic can be implemented iteratively, whereby newly incoming data from monitoring systems and control tests further improve the model and reduce uncertainty.

6.4. Integration of generative models with training programmes: a hybrid approach

The hybrid approach combines the global capabilities of generative models (VAE/GAN) to explore and synthesise the admissible space with traditional local optimisation methods that provide fine-tuning of parameters. Thus, the risk is reduced that local methods start from irrelevant initial points and become trapped in local optima.

A key element is the introduction of contextual and conditional variables (e.g., stage of the annual cycle, remaining days to a control/main competition), through which methodological admissibility and compliance with periodisation logic are ensured.

The dissertation examines major classes of generative models (VAE, GAN and diffusion models), and argues that, for tasks of generating training programmes, a key requirement is

reproducibility of distributions across training load zones and preservation of internal temporal logic. This places emphasis on conditional generation and on controlling inconsistencies through additional criteria and rules.

Characteristic problems in generative architectures are discussed, including training instability and “mode collapse” in GANs, which can lead to limited diversity and methodologically inadequate synthesis. The methodology incorporates procedures for controlling input data, stratification by $\text{VO}_{2\text{max}}$, generator–discriminator synchronisation and periodic validation on a separate set, which limits the risk of overfitting and increases training robustness.

Hybrid algorithms described in the scientific literature (“Global–Local Synergy”, “Tunnel Generative + Adaptive Search”, “Multi-Objective Generative Pipeline”) are also presented, in which the generative core creates a set of candidates and local optimisation procedures perform fine-tuning with respect to specific goals and constraints. In the thesis, these approaches are critically analysed and used as a basis for constructing a comprehensive integral methodology with clearly defined stages and verification criteria.

Training strategies under limited data are also considered, enabling adaptation of the model to specific profiles without loss of dependencies learned from a larger training sample.

7. AN INTEGRAL METHODOLOGY FOR OPTIMISING TRAINING LOAD THROUGH GENERATIVE ARTIFICIAL INTELLIGENCE MODELS

An integral methodology is understood as a system of purposeful means, methods and forms for optimising training load in cyclical sports on the basis of established principles through generative artificial intelligence models. The methodology is linear in nature and structured in a strict logical sequence, with each stage following completion of the activities in the preceding stage.

7.1. Stage 1: Designing the information and functional architecture

- Defining the objectives of the generative model and describing the expected outcomes through measurable metrics.
- Designing training datasets: planning volume and scope, database schema, relational links, indexing, metadata and version control.
- Designing outputs: the structure of personalised training programmes, presentation format, interface and visualisation solutions.
- Contextual description: data semantics, relationships between variables, usage scenarios, temporal dynamics.
- Selecting methods and tools for data collection (including expert-generated data) and defining the boundaries of the training and test samples.
- Designing test scenarios for working with the model, including standard and extreme conditions.

Stage 1 covers activities related to goal-setting, design of samples and datasets, description of metadata, selection of a data-collection methodology and selection of test scenarios, with the

aim of ensuring information and functional readiness for training, evaluation and application of the generative model.

7.1.1. Defining the objectives of the generative model

This sub-stage focuses on formulating specific objectives and expected outcomes from applying the generative model for optimising training load.

Goal-setting should be oriented towards the model's capacity to generate personalised training programmes on the basis of established principles, with results presented through clearly defined and measurable metrics that enable an objective assessment of programme effectiveness and functional effect.

7.1.2. Designing training datasets for model training

Designing datasets is associated with planning data volume and scope and preparing and structuring the data in a way that maximises their usefulness and efficiency in the training process.

- Planning data volume and scope: determining the required volume of training data and the scope of data to be included in the dataset, including time range and relevant parameters.
- Creating a database schema: defining relational links between different tables and datasets and selecting appropriate indices to accelerate access and processing.
- Defining metadata: creating documentation for each field (meaning, type, admissible values and relationships) and version control of data and schema to ensure traceability and recovery.
- Data integration and compatibility: standardising formatting and encoding and developing procedures for consolidating data from different sources into a unified dataset with guaranteed consistency and integrity.

7.1.3. Designing outputs

The model's ability to create personalised training programmes requires a clear structure and precision in the design of outputs, as well as the definition of measurable metrics for evaluating effectiveness.

- Defining personalised training programmes: creating programmes aligned with individual goals, physical capacities and the athlete's current condition.
- Interface for presenting results: developing informative visualisations of results (text, graphs, tables, interactive dashboards) to support expert interpretation and control.

7.1.4. Contextual description of the data

Contextual description of the data ensures semantic understanding and correct processing of input parameters and reduces the likelihood of generating irrelevant outputs.

- Data semantics: including contextual information on the substantive meaning of the data, definitions of measurement units, descriptions of data categories and relationships between variables.
- Relationships and dependencies: describing contextual dependencies between data, including the influence of factors on the training process and interrelations between indicators.

- Examples and scenarios: providing examples of real data and usage scenarios to support interpretation in different situations.
- Data dynamics: describing temporal aspects (update frequency, temporal dependencies) and dynamic relationships between variables.

7.1.5. Selecting methods and tools for data collection

This sub-stage includes selecting methods and tools for collecting the datasets required to train the generative model, with methods classified as data acquisition, collection, extraction and transfer.

- Data acquisition: web scraping; API integrations for access and extraction of data from external systems; sensors and IoT devices for real-time data collection.
- Data collection: surveys and questionnaires; observation; focus groups and interviews.
- Data extraction: data mining to extract hidden patterns and trends; text mining for analysing textual data; processing large volumes of data (Big Data); expert-generated data containing logically consistent and validated structures for training.
- Data transfer: file transfer (FTP/SFTP), cloud storage, database replication, data pipelines (message queues and data streams) for real-time exchange.

7.1.6. Defining the boundaries of the training and test samples

This sub-stage is directly related to model training and testing at a subsequent hierarchical level of the methodology and includes definition of the training and test samples.

- Defining the training sample: scope of data and diversity in the training data; representativeness of the sample for the general population to which the model applies; adequacy analysis to determine the minimum data volume that ensures sufficient accuracy.
- Defining the test sample: independence from the training data for objective assessment; representativeness; optimal size and statistical significance of results.

7.1.7. Designing test scenarios for working with the generative model

Designing test scenarios is an essential sub-stage for assessing the model's accuracy, reliability, adaptability and effectiveness under real-world conditions; scenarios should cover a wide range of situations.

- Defining the objectives of the test scenarios: specifying objectives for evaluating accuracy and parameterising the model under different scenarios.
- Developing test scenarios: scenarios for athletes with different levels of preparation; standard training conditions reflecting everyday practice; extreme training conditions to test robustness.

The development of test scenarios aims to provide an empirically grounded framework for examining the generative model's (GAN) ability to reproduce realistic training programmes reflecting regularities established in the training dataset. In this sense, scenario control is simultaneously a parameterisation tool (by setting input conditions) and a verification tool (by checking expected output relationships).

The test scenarios are designed to reflect different levels of aerobic capacity ($VO_2\text{max}$) and different temporal proximity to control and main competitions. The scenario set includes

VO₂max values in the range 50–80 ml·kg⁻¹·min⁻¹, which enables stratification of athletes into four quartile subgroups (Q1–Q4)—from low to high level of preparedness.

Temporal proximity to control competitions (Control_days) serves to check whether the model reproduces regularities in the days immediately preceding a control competition (t–1), as follows:

- absence of high-intensity regimes (Zone 4–Zone 5);
- absence of strength loads (Strength = 0);
- presence of compensatory low-intensity work in Zone 1.

Temporal distance to main competitions (Main_days) in the analysed programmes varies between 81 and 321 days, defining a long and very long period to the main start. This variable sets the seasonal context of load and supports distinction between base preparation and stages in which a higher relative share of specific work is expected.

Based on the combinations of VO₂max_input, Control_days and Main_days, five main test scenarios (S1–S5) are defined, simulating different types of training situations and imposing specific methodological constraints on the output of the generative model:

Scenario 1 (S1) – low VO₂max (Q1), Control_days = 1, Main_days ≥ 12 weeks. A scenario for poorly trained biathletes, in which complete absence of high-intensity loads (Zone 4–5), no strength work, and a short compensatory volume in Zone 1 are expected.

Scenario 2 (S2) – medium VO₂max (Q2–Q3), Control_days = 1, Main_days ≥ 12 weeks. A scenario for athletes with a moderate level of preparation, characterised by a moderate volume in Zone 1 in the absence of Zone 4–5 and Strength.

Scenario 3 (S3) – high VO₂max (Q4), Control_days = 1, Main_days ≥ 12 weeks. A scenario for highly trained biathletes, in which zero values in Zone 4–5 and Strength are expected, with a short priming session in Zone 1.

Scenario 4 (S4) – low VO₂max (Q1), Control_days = 1, Main_days >> 12 weeks. A scenario for the base preparatory stage in poorly trained biathletes, where day t–1 is recovery or low-intensity aerobic work without strength loads.

Scenario 5 (S5) – high VO₂max (Q4), Control_days = 1, Main_days >> 12 weeks. A scenario for the preparatory stage in highly trained biathletes, where controlled reduction of intensity before the control competition is expected while preserving a low-intensity priming component.

7.2. Stage 2: Aggregating training datasets

This stage includes identifying data sources, data types and access, as well as collecting and consolidating data into a unified dataset. Emphasis is placed on structured tabular data, standardisation of units and formats, and consistency control.

Within the methodology, Stage 2 is specified through sub-stages aimed at ensuring relevance, representativeness and technological reproducibility of the data for training the generative model.

- Identifying sources, data type and access: selecting relevant sources (internal databases, public repositories, application programming interfaces, etc.) and classifying data as structured and unstructured.
- Data extraction: extracting useful information from previously collected data, including data mining to identify patterns and trends in large datasets and expert-

generated data containing logically consistent, validated structures for training generative models.

- Data transfer: transferring data between systems (file transfer, cloud storage, replication, data streams) in order to consolidate into a unified dataset and ensure integrity and traceability of changes.
- Defining the boundaries of the training and test samples: defining scope, representativeness and sufficiency of the training sample, as well as independence and optimal size of the test sample for objective performance assessment.
- Designing test scenarios: developing scenarios to verify accuracy, reliability, adaptability and effectiveness of the model under different contexts and conditions of the training process.

7.3. Stage 3: Processing the datasets

- Cleaning the data of errors: identifying and correcting anomalies, removing duplicate records, handling missing values, standardising formats.
- Validating the data against pre-defined ranges and types; checking scope and constraints.
- Data storage: selecting infrastructure (SQL/NoSQL/cloud), security (access control, encryption), backup and recovery, traceability.

7.4. Stage 4: Managing the generative model

- Controlling input and output data through specific validation rules and procedures.
- Training the generative model (training with one/several examples; fine-tuning) in accordance with the specifics of the sport discipline and available data.
- Configuring the network for training: learning rate, number of epochs, validation during training; metrics (MAE, KL divergence, share of logical inconsistencies).
- Synchronising training (generator–discriminator) and managing risks such as mode collapse in GANs.

7.5. Stage 5: Evaluating model effectiveness

- Defining success criteria, minimum acceptable thresholds and a mechanism for expert assessment.
- Metrics of difference between observed and generated values (e.g., mean absolute error).
- Measures of similarity of distributions (KL divergence) and analysis of structural relationships.
- Checking logical and physiological coherence through formalised rules.
- Empirical validation through control experiments and feedback.

7.5.1. Formalised logical objectives and rules for parameterisation and verification

Within the methodology, objectives and rules derived from the empirical analysis of the training programmes are defined; they serve both as conditions for parameterising the GAN and as criteria for logical verification of generated results:

1. To ensure that input $\text{VO}_{2\text{max}}$ values fall within the interval $[50; 80] \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ and follow the real distribution with quartile boundaries $Q1=56, Q2=62, Q3=71$.
2. To ensure stratified generation by $\text{VO}_{2\text{max}}$ ($Q1-Q4$) in order to reduce internal dispersion and model more precisely the relationships between $\text{VO}_{2\text{max}}$ and load.
3. To reproduce the monotonic relationship between $\text{VO}_{2\text{max}}$ and load volume, generating higher median values of daily volume and of Zone 1–Zone 3 for higher quartile subgroups ($Q3-Q4$).
4. To preserve aerobic dominance, whereby Zone 1–2 account for $\geq 75\%$ of total daily volume, while Zone 4–5 and Strength are limited to $\leq 25\%$.
5. To reproduce the real shares of zero values for high-intensity and strength regimes within intervals compatible with the training sample (Zone 4; Zone 5; Strength).
6. To preserve the frequency of high-intensity days in the highest $\text{VO}_{2\text{max}}$ quartile ($Q4$), so that approximately half of days contain an intensive component ($Z4+Z5>0.1$).
7. To comply with the mutual-exclusion principle between high-intensity and strength loads, so that in $\geq 85\%$ of days with ($Z4+Z5>0.1$) the condition $\text{Strength} = 0$ is satisfied.
8. To prevent accumulation of excessive intensity by allowing no more than two consecutive days with ($Z4+Z5>0.1$), and by disallowing a high-intensity day followed immediately by a strength-training day.
9. To ensure internal logical coherence of the microcycle by preserving the sequence of load and recovery characteristic of real training programmes.

8. EMPIRICAL BASIS, DATA AND PARAMETERISATION OF THE GENERATIVE MODEL

Training and testing of the generative model are based on expert-designed training programmes for the youth national biathlon team, structured at the daily level. Access to the training programmes was provided in Excel (.xlsx) format as a structured, machine-readable dataset comprising 44,694 observations (daily training units) and 200 uniquely identified training programmes (Program_ID).

8.1. Input and output parameters

The input parameters used to personalise the generated training programmes are:

- $\text{VO}_{2\text{max_input}}$ – maximal oxygen uptake ($\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$), used to individualise load according to aerobic capacity;
- Control_days – number of days until the next control competition;
- Main_days – number of days until the next main competition.

The model output is a daily distribution of training volume across training load zones (Zone 1–Zone 5) and strength training (Strength), expressed in hours, as well as a day type (Type) indicating a training day or a control/main competition day.

8.1.1. Description of variables (daily record structure)

Table 2 presents the notation of observations used in the training and test samples. It integrates the athlete profile, calendar context and distribution of daily training volume across training load zones.

Each daily record in the dataset contains a combination of quantitative and qualitative indicators describing: (1) the athlete's physiological profile (VO₂max_input), (2) parameters related to participation in competitions (Control_days, Main_days, Control_dates, Main_dates), (3) the chronological positioning of the session (Day_index) and (4) the detailed distribution of training volume across training load zones (Zone 1–Zone 5) and strength training (Strength), as well as (5) the day type (Type) indicating a control or main competition.

The notation of the main variables in the daily record can be summarised as follows:

- Program_ID – unique identifier of a training programme, used to group daily records within the same programme;
- VO₂max_input – maximal oxygen uptake ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$) as an input parameter for individualising load according to aerobic capacity;
- Control_days / Main_days – remaining days to a control/main competition as contextual (conditional) variables;
- Control_dates / Main_dates – dates of control and main competitions as a calendar marker of the competitive context;
- Day_index – ordinal day of the week (1 = Monday, ..., 7 = Sunday), used for chronological ordering and microcycle analysis;
- Zone 1–Zone 5 – duration (in hours) of work in the respective training load zone, from low-intensity aerobic work (Zone 1) to maximal intensity (Zone 5);
- Strength – time (in hours) allocated to strength training;
- Type – marker indicating control and main competitions.

To model more effectively the relationships between the main components of observations and to ensure traceability at the level biathlete–programme–day–load, a relational data structure was applied, organised in logical tables (profile, programme, daily session and load). This organisation provides data normalisation, system scalability and the possibility of indexing by key fields (e.g., Program_ID and Day_index), facilitating training and validation of the generative model and enabling conditional queries on data subsets.

Table 2. Description of variables in daily training observations.

Variable	Description	Value type
Program_ID	Unique identifier of a training programme; used to group daily records within the same programme.	Text
VO2max_input	Input parameter – maximal oxygen uptake ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$), used to individualise load according to aerobic capacity.	Integer
Control_days	Number of days until a control competition; used to monitor form and adaptation.	Integer
Main_days	Number of days until a main competition; sets the target peak of sporting form.	Integer
Control_dates	Dates of control competitions/tests (e.g., 10	Date/Text

	km running, shooting tests).	
Main_dates	Dates of main competitions.	Date/Text
Day_index	Ordinal day of the week (1=Monday ... 7=Sunday) for chronological ordering and microcycle analysis.	Integer
Zone 1	Duration of training in Zone 1 (low-intensity aerobic work/recovery sessions), in hours.	Decimal number
Zone 2	Duration of training in Zone 2 (moderate aerobic work – basic endurance), in hours.	Decimal number
Zone 3	Duration of training in Zone 3 (mixed work – aerobic/anaerobic transition), in hours.	Decimal number
Zone 4	Duration of training in Zone 4 (high intensity around the anaerobic threshold), in hours.	Decimal number
Zone 5	Duration of training in Zone 5 (maximal intensity; short intervals/simulations), in hours.	Decimal number
Strength	Time allocated to strength training (strength, explosiveness, stability), in hours.	Decimal number
Type	Day type: training day / control competition / main competition.	Text

8.2. Stratification by VO₂max and median daily loads

To reduce internal variation and model more precisely the relationships between aerobic capacity and training loads, stratification by VO₂max quartile subgroups (Q1–Q4) was applied. Table 3 presents median daily loads (in hours) by training load zones and total volume. Variance analysis of the input variable VO₂max indicates a wide range (min = 50; max = 80 ml·kg⁻¹·min⁻¹) and moderate positive skewness ($Sk \approx 0.29$), which means concentration of observations at lower values and the presence of a right tail with less frequent high values. Negative kurtosis ($K \approx -1.133$) indicates a flatter distribution with thinner tails relative to the normal distribution.

Descriptive indicators for VO₂max are summarised via quantile estimates: Q1 = 56; median = 62; Q3 = 71 ml·kg⁻¹·min⁻¹, with a mean of 63.69 and population standard deviation $\sigma = 8.736$. This profile supports the use of non-parametric estimates as robust measures of variability and justifies stratification by quartile groups (Q1–Q4).

Stratification limits within-group dispersion and enables conditional parameterisation of the generative model, whereby relationships between aerobic capacity and training loads are

modelled more accurately for each level of preparedness. In practice, this means that generated daily loads are modulated according to membership of Q1–Q4, while preserving the empirically established tendency for higher median values of Zone 1–Zone 3 and total volume at higher VO₂max levels.

Table 3. Quartile distribution of median daily loads (by VO₂max)

VO ₂ max quartile	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Strength	Total load
Q1	43.9	16.1	3.2	0	0	0	75.7
Q2	46.8	17.8	3.2	0	0	0	80.7
Q3	52.3	19.6	3.4	0	0	0	89.4
Q4	59.3	20.8	3.7	0	0	0	93.8

Median values show a clear monotonic relationship between aerobic capacity level and load volume: as VO₂max increases, a gradual increase is observed in Zone 1–3 and total volume. High zones (Zone 4–5) and strength loads have values close to zero, reflecting dominance of low- and moderate-intensity work during base preparation.

Additional parameterisation is implemented through indicator relationships (conditional probabilities) describing the frequency of co-occurrence of high intensity and strength training in real programmes, and used as control constraints during generation.

- $P(Z4+Z5>0.1 \mid VO_{2\max} \in Q4) = 0.5087$ – the probability that the combined load in Zone 4 and Zone 5 exceeds 0.1 h (≈ 6 minutes) given that the athlete is in the highest VO₂max quartile (Q4). The value indicates that approximately 51% of days in this group contain a distinct high-intensity component.

- $P(\text{Strength} = 0 \mid Z4+Z5>0.1) = 0.8547$ – the probability that strength load is zero given the presence of a high-intensity load ($Z4+Z5>0.1$). The high value means that in about 85% of cases where high intensity is present, strength training is absent on the same day.

These relationships reflect the principle of functional incompatibility between strength and high-intensity loads within the microcycle and serve as a logical framework for verifying generated programmes.

8.3. Data access and technological implementation

In accordance with the methodology, access to the training programmes was implemented via an Excel (.xlsx) file format, which constitutes a structured, machine-readable dataset suitable for storing tabular data of heterogeneous type (numeric and text values).

The specific file structure includes: a structured table in rows and columns, 44,694 observations (daily training units), uniquely identified programmes (Program_ID) and structured variables describing the athlete profile, calendar context and distribution of load across training load zones (Zone 1–Zone 5) and the “Strength” module (Strength).

For the purposes of analysis and integration into the model architecture, the data were processed through scripting using the Python programming language and the pandas library, which provides functionality for direct loading of .xlsx files, conversion into tabular (DataFrame) format and subsequent transformation into structured inputs for the model.

The technological implementation complies with requirements for traceability and reproducibility of processing, with each transformation (validation of types, ranges and

consistency, standardisation of formats and units, aggregation by days and programmes) performed according to formalised rules.

8.4. Legal compliance and ethical rationale

The methodology considers, in an integrated manner, the legal and licensing status of the data, file formats and software tools used, aiming to ensure legality, transparency and reproducibility of the study.

In this context, the data were used as structured training programmes provided for scientific purposes, while processing and analysis were performed using widely used software tools and formats that ensure standardisation and access control.

9. GENERATIVE MODEL DESIGN AND VERIFICATION

9.1. Architecture selection and functional formulation

The dissertation substantiates the choice of a GAN architecture as a means of generating training programmes. The GAN model is treated as an instrument for global synthesis of training scenarios, in which the generator produces new daily loads and the discriminator distinguishes real from generated programmes. The adversarial mechanism supports reproduction of the statistical structure of the training data.

9.2. Control and standardisation of input and output data

Prior to training, input data control was implemented through validation of types, ranges and consistency between variables. Standardisation aims to limit the influence of differing parameter scales and to support convergence of training. Output control includes checking compliance with periodisation logic, alternation between load and recovery, and conformity with contextual conditions (Control_days, Main_days).

9.3. Metrics for assessing the quality of generated programmes

The quality of generated training loads was assessed through: (1) metrics of difference between observed and generated values (mean absolute error—MAE), (2) measures of similarity of distributions (KL divergence), and (3) checking logical and physiological coherence via formalised rules.

Mean absolute error (MAE) is used as an indicator of model accuracy in reproducing daily values (in hours/minutes) across energy regimes. For each load regime (Zone 1–Zone 5 and Strength), the mean absolute difference between observed and generated daily values is calculated.

Similarity of distributions between observed and generated programmes is evaluated via KL divergence (Kullback–Leibler), computed on normalised relative shares across training load zones:

The quantitative metrics (MAE and KL divergence) are complemented by formalised logical rules to control internal consistency and methodological admissibility (e.g., constraints on loads on days associated with control and main competitions, and constraints on the sequence of high-intensity days within the microcycle).

9.4. Analysis of structural similarity between training and generated programmes

Analysis of microcycle structure shows that approximately three quarters of the programmes in both the training and generated samples contain at least one sequence of three or more high-intensity days; however, the overall share of such days relative to all high-intensity days is only 3.8–4.2%. This indicates that intensive sessions are brief and isolated, in accordance with the principles of load and recovery in biathlon.

The comparison confirms that the model has learned not only quantitative relationships (distribution by zones) but also the functional structure of load, including alternation between high- and low-intensity work within microcycles.

Training verification was carried out using a separate validation set (approximately 10% of programmes), on which the dynamics of indicators were tracked after each epoch to prevent overfitting to the training data.

During training, an optimiser was used with parameters $\beta_1=0.5$ and $\beta_2=0.999$, with fixed learning rates $\eta_G=\eta_D=2\times 10^{-4}$.

The batch size was 512 daily records, providing sufficient gradient stochasticity and convergence stability. Training was conducted for 80 epochs at an update ratio $D:G = 1:1$, with shuffling of records in each epoch to maintain balance between generator and discriminator and to avoid discriminator saturation.

9.4.1. Checking logical consistency (rules over daily records)

To verify that the generated programmes follow the methodological logic of the training process, rules over daily records were defined and applied identically to the training and test samples.

Rule 1 (competition day): requirement for absence of load—Zone 1 = Zone 2 = Zone 3 = Zone 4 = Zone 5 = Strength = 0.

Rule 2 (consecutive high-intensity days): requirement that no more than two consecutive days with high intensity occur, where high intensity is defined as $\text{HighIntense} \equiv (\text{Zone 4} + \text{Zone 5}) > 6 \text{ min}$ (6 min is the minimum threshold for a high-intensity component).

All checks were performed in minutes, with the rules applied to daily records for the entire period.

Table 4. Results of the logical-consistency check.

Rule	Indicator	Training sample (n)	Test sample (n)
1. Competition day – Zone 1 = Zone 2 = Zone 3 = Zone 4 = Zone 5 = Strength = 0	Number of days	3490	401
	Cases with load	0	0
	Share with load (%)	0	0
2. Consecutive high-intensity days – no more than two consecutive days with Zone 4 + Zone 5 > 6 min	Number of programmes	180	20

	Programmes with a sequence ≥ 3 days	141	15
	Share of programmes with a sequence ≥ 3 days (%)	78,33	75
	Total high-intensity days	5333	574
	Days in sequence ≥ 3 days	225	22
	Share of all high-intensity days (%)	4,22	3,83

Full compliance with the rest rule is established. In neither dataset is load observed in any zone (0% cases with activity), indicating correct reproduction of the logic whereby a control or main competition day is not combined with training work.

Approximately three quarters of the programmes in both samples contain at least one sequence of three or more high-intensity days, but the overall share of such days relative to all high-intensity days is only 3.8–4.2%. This means that intensive sessions are brief and isolated, while the required alternation between load and recovery within the microcycle is preserved.

Comparison of results shows a high degree of similarity between the training sample and the generated programmes—both in the frequency of sequences and in the relative share of high-intensity days. The similarity confirms that the model learns not only quantitative relationships in the data but also the functional structure of load.

10. EXPERIMENTAL APPLICATION: YOUTH NATIONAL BIATHLON TEAM (SEASON 2023–2024)

To verify the effectiveness of the developed generative model, an experiment was conducted under real-world conditions, in which a training programme generated by the model was applied to athletes from the youth national biathlon team during the 2023/2024 period.

10.1. Control tests and indicators

For objective assessment, baseline and follow-up control tests were conducted under standardised conditions (identical course, temperature and surface state). Three key indicators were measured:

- Maximal oxygen uptake ($\text{VO}_{2\text{max}}$) – an indicator of aerobic capacity.
- Time for a 10 km roller-ski run (Marwe 6) – an integrative indicator of overall endurance, technical efficiency and specific speed endurance.
- Mean running speed ($\text{m}\cdot\text{s}^{-1}$) – derived from time and distance, characterising endurance.

The Marwe 6 roller-ski type is a standard in control tests for biathlon and cross-country skiing, as it provides high repeatability of results and mechanical resistance close to that of gliding on snow.

10.2. Individual results from baseline and follow-up tests

Table 5 presents the individual results of eight athletes (n=8), as well as the differences between the baseline and follow-up tests (Δ).

Table 5. Baseline and follow-up test results (10 km roller skiing, Marwe 6)

Biathlete	VO ₂ max baseline (ml·kg ⁻¹ ·min ⁻¹)	VO ₂ max follow-up (ml·kg ⁻¹ ·min ⁻¹)	10 km baseline (mm:s s)	10 km follow-up (mm:s s)	Δ time (s)	Speed baseline (m·s ⁻¹)	Speed follow-up (m·s ⁻¹)	Δ speed (m·s ⁻¹)
1	60.8	63.6	27:04	24:34	-150	6.16	6.79	0.63
2	65.0	67.8	28:46	27:05	-101	5.80	6.15	0.35
3	63.6	69.0	26:38	24:14	-144	6.26	6.88	0.62
4	60.0	65.4	26:49	25:14	-95	6.22	6.61	0.39
5	61.8	66.0	27:24	25:37	-107	6.08	6.51	0.43
6	66.3	69.9	28:59	26:41	-138	5.75	6.24	0.49
7	58.1	62.1	26:52	25:20	-92	6.21	6.58	0.37
8	55.0	59.0	30:30	29:10	-80	5.47	5.71	0.24

10.3. Variance and statistical analysis of effect

Variance analysis indicates high group homogeneity (CV < 7%) and physiologically consistent dynamics among the three variables: an increase in VO₂max is accompanied by a decrease in time and an increase in speed.

Table 6. Variance analysis of baseline and follow-up tests (n=8)

Indicator	n	Min	Max	R	Mean	SD	Skewness (Sk)	Kurtosis (K)	V (%)
VO ₂ max – baseline test (ml·kg ⁻¹ ·min ⁻¹)	8	55	66.3	11.3	61.33	3.69	-0.32	-0.75	6.01
VO ₂ max – follow-up test (ml·kg ⁻¹ ·min ⁻¹)	8	59	69.9	10.9	65.98	3.73	-0.24	-0.91	5.65
10 km – baseline test (s)	8	1598	1830	232	1698.9	82.7	0.52	-0.94	4.87
10 km – follow-up test (s)	8	1454	1750	296	1564.4	103.6	0.81	0.17	6.62
Speed – baseline test (m·s ⁻¹)	8	5.47	6.26	0.79	5.99	0.28	-0.29	-1.13	4.68
Speed – follow-up test (m·s ⁻¹)	8	5.71	6.88	1.17	6.43	0.40	-0.33	-1.01	6.22

To check assumptions, the Shapiro–Wilk test ($\alpha = 0.05$) was applied to the differences between baseline and follow-up tests. On this basis, the non-parametric Wilcoxon test for paired samples was selected.

Wilcoxon test results show statistically significant differences between baseline and follow-up tests for all three indicators, with p-value = 0.008 (< 0.05) for each. This means that the probability that the observed changes are due to chance is below 1%.

Effect size was calculated using $r = Z/\sqrt{n}$; the obtained value $r = 0.94$ ($n = 8$) indicates a large effect (according to Cohen’s classification) and a very strong practical impact of the applied training programme on the athletes’ functional indicators.

For completeness, the results of the normality check (Shapiro–Wilk) and the significance test (Wilcoxon) are presented in Table 7 and Table 8.

Table 7. Shapiro–Wilk test results for normality of differences ($n=8$; $\alpha=0.05$).

Indicator (follow-up minus baseline)	p-value	Interpretation ($\alpha=0.05$)
VO ₂ max (ml·kg ⁻¹ ·min ⁻¹)	0.047	The distribution differs from normal ($p<0.05$)
10 km time (s)	0.031	The distribution differs from normal ($p<0.05$)
Speed (m·s ⁻¹)	0.039	The distribution differs from normal ($p<0.05$)

Table 8. Wilcoxon test results for paired samples ($n=8$; $\alpha=0.05$).

Indicator	p-value	Interpretation ($\alpha=0.05$)
VO ₂ max (ml·kg ⁻¹ ·min ⁻¹)	0.008	Statistically significant difference – increase in VO ₂ max
10 km time (s)	0.008	Statistically significant difference – decrease in time
Speed (m·s ⁻¹)	0.008	Statistically significant difference – increase in speed

The planned training load for season 2023–2024 covers the period 01/04/2023–31/03/2024; for each day, minutes of work in five training load zones and time allocated to strength training are recorded, as well as period affiliation within the preparation cycle.

The total volume of work performed over the period amounts to 34,045 minutes (567.4 hours). Of these, 30,695 minutes (511.6 hours) are endurance work in Zones 1–5, and 3,350 minutes (55.8 hours) are allocated to strength training. Strength training accounts for 9.8% of total annual volume and 10.9% of endurance training volume.

The distribution of time across training load zones shows a clearly pronounced “pyramidal” pattern: Zone 1 – 293.9 hours (57.4% of time in Zones 1–5); Zone 2 – 143.8 hours (28.1%); Zone 3 – 43.4 hours (8.5%); Zone 4 – 18.4 hours (3.6%); Zone 5 – 12.1 hours (2.4%). In total, Zone 1 and Zone 2 account for 85.5% of endurance training time, whereas Zones 3–5 account for 14.5%, which is typical for endurance sports and compatible with principles of periodisation in biathlon.

11. RESULTS AND ANALYSIS

The dissertation results are grouped into two main directions: (1) methodological results—development and formalisation of an integral methodology and model architecture for optimising training loads through generative models; (2) empirical results—statistical and practical verification of the model through analysis of similarity between training and generated programmes and experimental application in a real training environment.

11.1. Results from analysis of zonal distribution and load logic

Analysis of the distribution of daily loads across zones (Zone 1–5) shows a regular dominance of low- and moderate-intensity loads (Zone 1–2), which corresponds to requirements for building an aerobic base and resilience in biathlon.

Stratification by $\text{VO}_{2\text{max}}$ demonstrates a monotonic relationship between aerobic capacity and load volume.

Daily and weekly load structure indicates 62 days without training load (16.9% of all calendar days), i.e., 304 training days, or on average about 5.8 training days per week. The mean daily training duration for the season is 93.0 minutes with a standard deviation of 63.5 minutes, a minimum of 0 minutes and a maximum daily value of 274 minutes.

The median daily load is 90.0 minutes, and the interquartile range (25th–75th percentile) is 53.5–135.0 minutes. Forty days with load ≥ 180 minutes were recorded (about 11% of all days).

The calculated mean weekly work volume is about 10.9 hours, characterising the load as relatively high, yet moderate and aligned with the age and developmental stage of youth national biathletes.

11.2. Strength training in the annual cycle

The realised loads show that strength training is concentrated in the preparatory period, where it accounts for about 10–11% of total load, whereas in the competitive period its share decreases to about 9%, and in the transition period it is virtually absent. The frequency of strength sessions is approximately 1.8 times per week in the preparatory period and 1 time per week in the competitive period. Strength training is conducted mainly on days with endurance work, which differentiates days with higher total load from days with lower total volume and requires precise control to limit the risk of overreaching.

The total realised strength volume over the study period is 3,350 minutes, or 55.8 hours, corresponding to 9.8% of total training load and 10.9% relative to the volume of work in Zones 1–5.

Strength training is recorded on 77 days, i.e., on average about 1.5 strength sessions per week across the season, with a mean duration of 43.5 minutes per strength session (median 45 minutes). The largest share of this volume is realised in the preparatory period—2,460 minutes, or 41.0 hours (73.4% of annual strength volume). Within the preparatory period, strength training accounts for 10.5% of total training load and 11.7% relative to work in training load zones 1–5. Fifty-seven strength-training days are recorded, i.e., approximately 2 strength sessions per week, with a mean duration of 43.2 minutes.

In the competitive period, realised strength volume amounts to 875 minutes (14.6 hours), or 26.1% of annual volume for this indicator. The share of strength within total load for the period is 8.8%, and 9.7% relative to work in Zones 1–5. Nineteen days with strength training

are recorded, corresponding to about 1 strength session per week, with a mean duration of 46.1 minutes. In the transition period, strength training is virtually absent—15 minutes (0.25 hours) are realised on a single day.

To evaluate how strength training is combined with endurance training within individual days, all training-load days were analysed (a total of 304 days), of which 77 days include strength training and 227 days include only work in training load zones.

On days with strength training, the mean volume of work in training load zones is 98.6 minutes (median 92.0 min; SD \approx 48.2 min), and strength training averages 43.5 minutes (median 45.0 min), forming a total daily load of 142.1 minutes. On days without strength training but with training, the mean value of work in training load zones is 101.8 minutes (median 92.0 min; SD \approx 46.9 min). Zone-work volume is practically identical on days with and without strength training; the substantive difference is observed in total daily load, which is approximately 40 minutes higher on days with strength training.

Considering only days with strength training, Pearson's correlation coefficient between minutes of strength work and minutes of work in training load zones is $r = 0.39$ (moderate positive correlation). When all training days are included, including those without strength training, the correlation between daily zone work and minutes of strength training practically disappears ($r \approx 0.05$) due to the large number of days with 0 minutes of strength alongside a wide range of zonal values.

On a weekly basis, the distribution of strength training by periods is approximately: about 1.3 hours of strength training and about 11.0 hours of zonal work per week in the preparatory period; about 0.8 hours of strength training and about 8.6 hours of zonal work per week in the competitive period; and a minimal volume of strength work alongside about 3.9 hours of weekly zonal work in the transition period.

These values confirm a clearly pronounced concentration of strength training in the preparatory period, with moderate presence in the competitive period and virtually no presence in the transition period. In microcycle terms, strength training is implemented predominantly on days with dominant endurance work, without a substantial reduction in zonal minutes, resulting in clear differentiation of days with higher total load and days with lower volume as part of the load–recovery alternation.

11.5. Analysis of annual volume and periodisation (season 2023–2024)

For descriptive analysis, data were grouped by training periods (preparatory, competitive and transition). Total work volume (in minutes and hours), the relative share of time in each training load zone, mean daily values and standard deviation, and the frequency of days without recorded training load were calculated.

Planned training load for season 2023–2024 covers the period 01/04/2023–31/03/2024; for each day, minutes of work in five training load zones (Zone 1–Zone 5) and time allocated to strength training (Strength) are recorded, along with affiliation to the transition, preparatory or competitive period.

The total volume of work performed over the period amounts to 34,045 minutes (567.4 hours). Of these, 30,695 minutes (511.6 hours) are endurance work in Zones 1–5, and 3,350 minutes (55.8 hours) are allocated to strength training. Strength training represents 9.8% of total annual volume and 10.9% relative to endurance training volume.

The distribution of time across training load zones follows a clearly pronounced “pyramidal” pattern: Zone 1 – 293.9 hours (57.4% of time in Zones 1–5); Zone 2 – 143.8 hours (28.1%); Zone 3 – 43.4 hours (8.5%); Zone 4 – 18.4 hours (3.6%); Zone 5 – 12.1 hours (2.4%). In total, Zone 1 and Zone 2 account for 85.5% of endurance training time, whereas Zone 3–Zone 5 account for 14.5%.

Daily and weekly load structure shows 62 days without training load (16.9% of all calendar days), i.e., 304 training days, or on average about 5.8 training days per week. The mean daily training duration for the season is 93.0 minutes with a standard deviation of 63.5 minutes; the minimum is 0 minutes and the maximum reaches 274 minutes. The median daily load is 90.0 minutes, and the interquartile range (25th–75th percentile) is 53.5–135.0 minutes. Forty days with load ≥ 180 minutes were recorded (about 11% of all days). The calculated mean weekly work volume is about 10.9 hours.

These indicators characterise the load as relatively high, yet moderate and aligned with the athletes’ age and developmental stage, with dominance of low- and moderate-intensity work and a controlled presence of higher zones in accordance with principles of sports periodisation.

The preparatory period spans 222 days and includes 194 training days, of which 28 are without training (12.6%). This period concentrates the largest share of annual load: total work volume 390.1 hours (68.8% of seasonal volume), endurance training volume (Zones 1–5) 349.1 hours, and strength training 41.0 hours (73.4% of annual strength volume).

The mean daily load duration in the preparatory period is 105.4 minutes with a standard deviation of 65.6 minutes; mean weekly volume is about 12.3 hours, and the mean duration of active training days is 120.6 minutes.

Zonal distribution in the preparatory period is characterised by dominance of Zone 1 and Zone 2 and a supportive share of high-intensity work: Zone 1 – 200.1 hours (57.3%); Zone 2 – 101.3 hours (29.0%); Zone 3 – 33.5 hours (9.6%); Zone 4 – 9.0 hours (2.6%); Zone 5 – 5.2 hours (1.5%). Zones 1–2 comprise 86.3% of endurance volume, and Zones 3–5 comprise 13.7%, with Zones 4 and 5 accounting for 4.1%.

The competitive period includes 123 days, of which 98 are training days and 25 are without training (20.3%). The realised volumes are: total work volume 165.5 hours (29.2% of annual volume), endurance training volume 150.9 hours, and strength training 14.6 hours (26.2% of annual strength volume).

The mean daily duration in the competitive period is 80.7 minutes with a standard deviation of 55.0 minutes; mean weekly volume is about 9.4 hours, and the mean duration of active training days is 101.3 minutes. Compared with the preparatory period, a reduction in total volume and an increase in the share of rest days are observed.

Zonal distribution in the competitive period is: Zone 1 – 87.9 hours (58.2%); Zone 2 – 37.9 hours (25.1%); Zone 3 – 9.3 hours (6.2%); Zone 4 – 8.9 hours (5.9%); Zone 5 – 7.0 hours (4.6%). The share of the high zones 4 and 5 increases to 10.5% (versus 4.1% in the preparatory period) at lower total volume, which is consistent with the logic of competition preparation.

The transition period includes 21 days, of which 12 are training days and 9 are without load (42.9%). Total load volume is 11.8 hours (2.1% of annual volume), with endurance work 11.6 hours and strength training 0.2 hours (12 minutes). The mean daily value is 33.7 minutes with a standard deviation of 35.0 minutes; mean weekly volume is about 3.9 hours, and the mean duration of active training days is 59.0 minutes.

Zonal distribution in the transition period is: Zone 1 – 5.9 hours (50.9%); Zone 2 – 4.6 hours (40.0%); Zone 3 – 0.6 hours (5.3%); Zone 4 – 0.4 hours (3.8%); Zone 5 – 0 hours, confirming the period’s character as “active recovery” with dominance of the lowest zones.

In summary, the descriptive analysis indicates a clearly structured annual periodisation: the preparatory period concentrates about two thirds of annual volume at predominantly low to moderate intensity; the competitive period is characterised by lower total volume and a relative increase in the high zones; the transition period is a recovery phase with minimal volume, dominance of Zones 1–2 and practical absence of Zone 5 work and strength training.

Table 9. Summary of annual work volume by training periods (season 2023–2024).

Period	Calendar days	Training days	Days without training (%)	Total volume (hours)	Endurance / Strength (hours)
Preparatory	222	194	28 (12,6%)	390,1	349,1 / 41,0
Competitive	123	98	25 (20,3%)	165,5	150,9 / 14,6
Transition	21	12	9 (42,9%)	11,8	11,6 / 0,2

Results in Table 9 show concentration of work volume in the preparatory period and a regular reduction of load in the competitive and transition periods, alongside an increase in the relative share of high zones in the competitive phase.

11.6. KL divergence and similarity of zonal distributions

To assess similarity between observed and generated zonal distributions, KL divergence (Kullback–Leibler) was calculated by VO₂max quartile groups. In each subgroup, the total volumes by zones (Zone 1–Zone 5 and Strength, in minutes) were normalised as relative shares of total daily volume and compared between observed (P) and generated (Q) data.

The obtained KL(P||Q) values are on the order of 0.0001–0.0004 in all quartile groups, indicating an extremely close shape of zonal distributions between observed and generated data. Similarity is uniform across Q1–Q4, with no quartile exhibiting a distinctly higher KL, which indicates stable conditional behaviour of the model with respect to VO₂max.

Table 10. KL divergence between observed and generated distributions by VO₂max quartiles.

VO ₂ quartile	Q1	Q2	Q3	Q4
KL(P Q)	0,0003	0	0	0,0001

Additional interpretation of similarity is presented through relative zonal shares (P – observed, Q – generated); below are included rows presented in the dissertation as part of the full set of values.

Table 11. Relative zonal shares (P – observed, Q – generated) (only a subset of all rows is shown).

VO ₂ quart.	Zone	P(observed)	Q(generated)
Q1	Zone 1	0.5482	0.5513
Q2	Zone 1	0.5473	0.5459
Q3	Zone 1	0.5484	0.539
Q4	Zone 1	0.5474	0.5523

Q1	Zone 2	0.244	0.2473
Q2	Zone 2	0.2432	0.2423
Q3	Zone 2	0.2449	0.2458
Q4	Zone 2	0.2447	0.2427
Q1	Zone 3	0.0752	0.0753
Q2	Zone 3	0.0765	0.0769
Q3	Zone 3	0.0758	0.0776
Q4	Zone 3	0.0775	0.0774
Q1	Zone 4	0.0261	0.0235
Q2	Zone 4	0.026	0.0269
Q3	Zone 4	0.0252	0.028
Q4	Zone 4	0.0245	0.0236
Q1	Zone 5	0.0184	0.0165
Q2	Zone 5	0.0178	0.0192

From the aggregated results it follows that the model correctly allocates total daily volume across training load zones within each level of preparedness. The smallest differences are observed in Zone 3 and Zone 5, whereas in Zone 1–2 deviations are minimal (within 0.002–0.010), which is expected given their dominant share.

11.7. Logical consistency

Logical verification complements quantitative metrics through formalised rules for controlling internal consistency and methodological admissibility of generated loads. Constraints over daily records were checked, related to absence of training load on a competition day and limiting consecutive high-intensity days under the definition $\text{HighIntense} \equiv (\text{Zone 4} + \text{Zone 5}) > 6 \text{ min}$. Verification results show full compliance with the rest rule (0% cases with activity on a competition day) and high structural robustness with respect to alternation between load and recovery within the microcycle. Although a substantial portion of programmes contain at least one sequence of three or more high-intensity days, the relative share of such days among all high-intensity days remains small (3.8–4.2%), characterising intensive loads as brief and isolated.

11.8. Summary of statistical verification

The combination of variance and quantile analysis, measures of distribution similarity (KL divergence), error metrics (MAE), tests of normality (Shapiro–Wilk) and significance (Wilcoxon), and evaluation of effect size (r) provides a complete quantitative and structural assessment of results and is methodologically adequate for the study data.

Table 12 and Table 13 present key quantitative indicators of the generative model's accuracy in reproducing the training data via mean absolute error (MAE), computed in minutes across energy regimes and by VO_2max quartile groups (Q1–Q4).

Table 12. MAE by training load zones (in minutes) and VO_2max quartile groups.

VO_2 quartile	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Strength	Mean MAE
Q1 (low preparedness)	1.09	0.64	0.12	0.17	0.12	0.02	0.36

Q2 (moderate preparedness)	1.46	0.67	0.15	0.01	0.08	0.24	0.43
Q3 (high preparedness)	0.51	0.72	0.37	0.35	0.20	0.46	0.44
Q4 (very high preparedness)	0.49	0.23	0.02	0.10	0.11	0.10	0.18

Table 13. Aggregated MAE values by training load zones (in minutes).

Training load zone	Mean MAE (min)
Zone 1	0.89
Zone 2	0.57
Zone 3	0.16
Zone 4	0.16
Zone 5	0.13
Strength	0.20
Mean overall MAE	0.35

Results indicate high accuracy in reproducing quantitative relationships between training regimes and level of preparedness. The mean overall MAE is 0.35 minutes, corresponding to below 1% deviation relative to mean load values in the training population. At the same time, the low MAE values for high-intensity zones and strength training confirm that the model reproduces rarer load components without disrupting microcycle logic.

11.4. Practical applicability

The experimental application of a generated training programme for season 2023–2024 demonstrates a real training effect, expressed in a statistically significant improvement in $\text{VO}_{2\text{max}}$, a reduction in 10 km test time and an increase in mean speed. This confirms the model’s operational effectiveness and its potential for implementation as a tool for analysing, planning and forecasting training load.

12. CONCLUSION

Integrating generative artificial intelligence models with contemporary sports theory constitutes a methodologically coherent and empirically validated approach to optimising the training process in cyclical sports.

- The training process can be reliably formalised as a multi-criteria optimisation problem in which the objective function reflects the adaptive effect and the control parameters are volume, intensity and recovery time.
- The proposed integral methodology of interrelated stages (design, aggregation, processing, management, evaluation) functions as a “closed-loop” management system in which data continuously refine the model and ensure reproducibility and traceability.
- The developed generative model reproduces quantitative and structural relationships of the training process and generates programmes comparable in structure and dynamics to those created by experts.

- Experimental application in the youth national biathlon team shows a statistically significant improvement in functional indicators and a very large practical effect, confirming the model's applicability in a real sports environment.
- Optimisation should be multi-criteria with a clearly defined vector of objectives; it is appropriate for classical optimisation methods to be used for local optimisation and for generative models—for global synthesis and selection of training scenarios.
- Periodic updating of parameters is recommended at the end of micro- and mesocycles on the basis of newly available data (fatigue, recovery indicators, etc.), as well as expansion of input data towards multimodal sources (physiological, biomechanical and psychological indicators).

The concept developed in the dissertation shows that the generative approach enables training programmes to be represented as multidimensional structures whose logic can be reproduced through learning from expert-designed annual plans. In this model, training load is considered as a controllable object, and the control parameters are described by volume, intensity, temporal structure and constraints arising from periodisation.

Empirical verification confirms that the quality of generated programmes can be assessed reliably through a combination of quantitative and logical criteria. Low values of mean absolute error (MAE) across energy regimes, extremely small KL divergence values between observed and generated zonal distributions, and results of the logical check of microcycle structure indicate that the model reproduces both quantitative relationships (zonal distribution) and the methodological sequence of load–recovery.

In the context of the youth national biathlon team (season 2023–2024), the practical applicability of the proposed approach is supported by results from baseline and follow-up control tests as well as statistical analysis of effect. Improvements in VO_2max , reductions in 10 km test time and increases in mean speed reflect the presence of a real training effect and show that the generated programme is not only statistically plausible but also functionally adequate in a real sports environment.

From a methodological perspective, a key advantage of the integral methodology is that it functions as a “closed-loop” management system: data from real preparation feed the generative model, results are evaluated by formalised criteria, and parameters are updated periodically at the end of micro- and mesocycles. This logic enables adaptive solutions under changes in functional state, calendar context and preparation objectives.

Practical implementation of the approach requires compliance with two mandatory principles. First, the generative model should be constrained by methodological rules that prevent physiologically inadmissible load combinations. Second, evaluation of results should be multi-criteria and should include both quantitative metrics (MAE, KL divergence) and logical coherence of microcycle structure and control over load content on specific days (e.g., before a control competition).

A limitation of the study arises from the fact that generative training reproduces regularities present in the training dataset. Therefore, the model's accuracy and robustness depend directly on the representativeness, quality and consistency of the training programmes used for training, as well as on the scope of input variables.

Prospects for methodological development include: expanding input data towards multimodal sources (physiological, biochemical and functional indicators), integrating fatigue and recovery

indicators, and extending logical rules to control methodological admissibility under a wider diversity of training means and competitive scenarios.

In summary, the dissertation shows that Generative AI can be used not as an autonomous replacement of the coach, but as a tool to support load planning and management through systematic generation of alternatives, evaluation of their quality and formalised alignment with the methodological principles of sports training.

13. SCIENTIFIC AND APPLIED CONTRIBUTIONS

13.1. Scientific contributions

1. An integral methodology has been developed for optimising the training process through generative artificial intelligence models, presenting a consistent stage structure—model design, data collection and processing, management of the generative algorithm and evaluation of effectiveness—which defines a systems approach to applying AI technologies in sports preparation.

2. The use of a GAN architecture as a means of generating training programmes has been substantiated.

13.2. Applied contributions

3. The possibility has been demonstrated for a generative artificial intelligence model to reproduce training programmes with minimal deviations relative to the real programmes used to train the model.

4. A system of formalised logical rules has been applied to control internal consistency and methodological admissibility of model-generated training loads.

5. A training programme for season 2023–2024, based entirely on a generative model, has been developed and applied for the purposes of the youth national biathlon team.