

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»**  
( Н И У « Б е л Г У » )

ИНСТИТУТ ИНЖЕНЕРНЫХ ТЕХНОЛОГИЙ И ЕСТЕСТВЕННЫХ НАУК  
КАФЕДРА ПРИКЛАДНОЙ ИНФОРМАТИКИ И ИНФОРМАЦИОННЫХ  
ТЕХНОЛОГИЙ

**РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ДЛЯ РЕШЕНИЯ ЗАДАЧ  
СЕЛЕКЦИИ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ**

Выпускная квалификационная работа  
обучающегося по направлению подготовки 09.04.03 Прикладная  
информатика очной формы обучения, группы 07001633  
Постникова Александра Николаевича

Научный руководитель  
д.т.н., доцент Черноморец А.А.

Рецензент  
к.ф.-м.н. О. В. Гальцев

БЕЛГОРОД 2018

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
1 Исследование современных подходов и информационных средств моделирования задач селекции.....	11
1.1 Обзор существующего информационного обеспечения моделирования задач селекции.....	11
1.2 Теоретические основы моделирования задач селекции .....	18
1.3 Обзор структур и методов обучения нейронных сетей .....	27
1.4 Анализ эффективности применения нейронных сетей.....	39
2 Разработка методов и алгоритмов моделирования процесса селекции ...	43
2.1 Разработка методов моделирования селекции.....	43
2.2 Проектирование модели решения задач селекции .....	55
2.3 Разработка программного интерфейса .....	58
2.4 Основные результаты и выводы.....	60
3 Проведение вычислительных экспериментов.....	61
3.1 Оценка эффективности разработанных методов.....	61
3.2 Основные результаты и выводы.....	69
ЗАКЛЮЧЕНИЕ .....	70
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	73

## ВВЕДЕНИЕ

В настоящее время информационные технологии широко используются во всех сферах деятельности человека, в том числе в сельском хозяйстве, включая растениеводство. Успехи в развитии растениеводства во многих случаях связано с наличием различных сортов растений способных выживать в конкретных климатических условиях и давать высокий урожай.

Основные направления для обеспечения сортового разнообразия — это селекция и эволюция. Поскольку эволюция процесс длительный, то в настоящее время применяется селекция, так как с ее помощью, под контролем специалистов, можно в ускоренном, по сравнению с эволюционным развитием, получать новые сорта. В данной работе рассматривается применение вычислительных технологий и искусственных нейронных сетей (ИНС), с помощью которых имеется возможность моделировать процесс селекции, а именно моделировать набор качеств будущих поколений от родительских особей. ИНС позволяет моделировать мутации генов и анализировать полученные результаты.

Актуальность работы обусловлена возможностью применения новых вычислительных технологий в селекции для моделирования процессе скрещивания и анализа результатов, внесения корректировок в исходные данные. Применение компьютерных технологий позволяет быстрее моделировать процессы селекции, что позволяет ускорить трудоемкий и длительный процесс отбора родительских особей и выведения новых с требуемым набором признаков. Наличие генетического разнообразия сортов растений, животных, лесов необходимо с целью сохранения популяционного разнообразия. В данной работе рассматривается возможность решения задач селекции с использованием средств вычислительной техники для выведения новых сортов растений.

Большое разнообразие сортов способствует пополнению продовольственных запасов на достаточном уровне, а повешенная выживаемость и урожайность позволяет собирать высокие объемы урожая.

Увеличить количество и качество собираемого урожая помогает селекция растений. Применение современных вычислительных технологий в области селекции дает возможность моделировать поведения процесса скрещивания растений, рассчитать выживаемость растений т.д. Возможность применения компьютерных средств обработки данных обусловлена их большим объемом и высокой скоростью работы с информацией, а также возможностью решения специфических задач селекции благодаря современным методам поиска, сбора, хранения, обработки и распространения информации.

Развитие вычислительной техники способствовало совершенствованию методов моделирования процессов поведения животных при помещении их в различные условия, а также ведение учета генов, отвечающие за необходимые свойства у объекта.

Селекция – это отрасль науки, являющийся основным источником развития Российской Федерации от успешного развития, которых зависит доля обеспеченности продуктами питания населения и развития экономики страны. Селекцию можно разделить на два основных сегмента: биологическая (эволюция), выведение новых сортов человеком. Особенностью селекции является возможность получения новых качеств продукта для адаптации к изменяющимся условиям окружающей среды. Человек использует селекцию для привития необходимых качеств.

На данный момент известные различные алгоритмы селекции, в разработке которых принимали участие В. С. Пустовойт, В. Н. Мамонтова, А. П. Шехурдин и другие. Ими проделана большая работа по увеличению урожайности отдельных представителей растительных культур.

Возможность применения в данной работе современных методов решения задач селекции обусловлена развитием технологий искусственного

интеллекта и искусственных нейронных сетей (ИНС). Использование ИНС в задачах селекции позволяет быстрее моделировать процесс. Анализ полученных результатов дает возможность корректировать входные данные.

Искусственные нейронные сети – это частичная реализация искусственного интеллекта, которая моделирует мыслительную деятельность человека.

Нейронные сети применимы для широкого спектра задач, в том числе и в процессе селекции. Задачи селекции подразумевают проведение экспериментов по скрещиванию нескольких растений для получения требуемых качеств у новых особей и занимает много времени.

Селекция представляет собой искусственное ускорение процессов эволюции. Проводя селекционную деятельность, специалистами были разработаны методы, которые адаптированы под разные условия работы, но направлены на достижение одной и той же цели.

Для эффективной работы селекционной промышленности необходимо проводить постоянные исследования, улучшать уже имеющиеся методы и алгоритмы с целью минимизации ошибки. Для этого необходимы практические исследования или моделирование процесса селекции на вычислительных машинах (ЭВМ). Использование ЭВМ снижает временные затраты, позволяя детально изучить процесс. Положительной стороной использования ИНС является гибкость и универсальность полученных методов.

Разработкой моделей селекции занимались ученые Н. И. Вавилов, Чарльз Р. Х., Д. Бродбента, Г.Д. Карпеченко и др. Существенный вклад в развитие моделей селекции в растениеводстве внес И.В. Мичурин. Исследователи в области селекции изучали следующие факторы: исходное сортовое и видовое разнообразие растений и животных; наследственную изменчивость; роль среды в развитии и проявлении нужных селекционеру признаков.

В настоящее время известны различные программные продукты, позволяющие решать отдельные задачи селекции. Для оценки селекционного потенциала разработано программное обеспечение BLUP. Известна программа по оцениванию питательных веществ в кормовой смеси растений. Программа STAGES позволяет проводить оценку различий старого поколения от нового. Программа Quantitative Trait Loci позволяет моделировать схему расположения генов. Программа "DIAS", разработанная Р.А. Цильке и Л.П. Присяжной, предназначена для расчёта генетических параметров различных сортов.

Анализ методов селекции и существующих программных продуктов показал высокий уровень теоретической проработанности аспекта моделирования задач селекции, и незначительное количество реализаций в виде программного обеспечения, в том числе использующего нейронные сети при моделировании процессов селекции.

Целью работы является разработка методов и алгоритмов для решения задач селекции с использованием искусственных нейронных сетей и генетического алгоритма, обеспечивающих моделирование процессов скрещивания родительских особей.

Для достижения поставленной цели решены следующие задачи:

- 1) Анализ методов и программных средств моделирования процессов селекции.
- 2) Разработка метода и алгоритма моделирования процесса скрещивания родительских особей с обменом генов.
- 3) Разработка метода и алгоритма моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети, обучаемой с помощью генетического алгоритма.
- 4) Разработка прототипа программной реализации разработанных алгоритмов.
- 5) Проведение вычислительных экспериментов по оцениванию работоспособности разработанных методов.

Объектом исследования являются информационные технологии, применяемые при моделировании процессов скрещивания растений, используемых в селекции.

Предмет исследования: моделирование процессов селекции на основе искусственных нейронных сетей.

В ходе исследования применялась методология системного анализа; использовались методы математической логики в представлении генетических моделей задач селекции.

Новизна работы состоит в разработке метода моделирования процесса скрещивания родительских особей с обменом генов, а также в разработке метода моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети, которые позволят проводить моделирование процесса скрещивания селекции на основе одного поколения. Параметры процесса мутации генов определяются на основе показателей генетического алгоритма.

Практическая значимость работы обусловлена разработанными алгоритмами моделирования процессов скрещивания родительских особей, а также созданным прототипом программной реализации разработанных алгоритмов, прошедшего апробацию на базе Управления сельского хозяйства Волоконовского района Белгородской области. Полученные данные апробации показывают работоспособность разработанных методов, алгоритмов.

Положения, выносимые на защиту:

- 1) Алгоритм моделирования процесса скрещивания родительских особей с обменом генов.
- 2) Алгоритм моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети, обучающейся с помощью генетического алгоритма.
- 3) Результаты проведенных вычислительных экспериментов, иллюстрирующие преимущество разработанных методов и алгоритма.

Автором осуществлена разработка методов моделирования задачи скрещивания родительских особей, алгоритмов, реализующих данные методы, прототипа их программной реализации, проведены вычислительные эксперименты.

В ходе исследования применялась методология системного анализа; использовались методы математической логики в представлении генетических моделей задач селекции, методическое программирование методов решения задач по селекции, анализ полученных результатов обучения полученной ИНС на основе разработанного алгоритма скрещивания, методы объектно-ориентированного программирования.

Магистерская работа состоит из введения, основной части, которая состоит из четырех разделов, заключения и списка используемой литературы.

Во введении сформулированы актуальность, цель и задачи работы, указаны научная новизна, положения, выносимые на защиту, практическая значимость работы, также приведено краткое содержание работы.

В первом разделе приведен обзор основных определений и комплекса средств обеспечения информационных технологий. Выделены задачи информационных технологий для решения задач селекции. Описаны существующие методы моделирования процессов селекции, которые используют различные методы проведения процедуры селекции. Приведена математическая модель работы известного метода моделирования селекции BLUP, а также результаты анализа применения данного метода.

Также, в первом разделе, проведен анализ методов селекции в природе. При обзоре методов были рассмотрены результаты применения их на практике. Выделены пути развития селекции и задач селекции, рассмотрены труды исследователей в изучаемой научной области.

В первом разделе также проведен анализ применения искусственных нейронных сетей, их структуры построения. Рассмотрены история возникновения и развития нейронных сетей. Проведен анализ известных

методов обучения нейронной сети, которая выявила положительные и отрицательны стороны различных подходов.

Приведены примеры применения нейронных сетей. Из приведенных примеров использования нейронных сетей определены положительные и отрицательные аспекты их применения.

Из анализа методов обучения был выбран генетических алгоритм, в качестве метода обучения без учителя, который минимизирует вычислительный процесс и не требует специалиста при первичном обучении нейронной сети.

Во втором разделе производится проектирование методов моделирования процессов селекции и разработка методов моделирования. Разработка алгоритма для метода моделирования процессов селекции с использованием нейронной сети и метода моделирования процессов селекции без использования нейронной сети, так же разработка метода моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети на первом этапе необходимо обучение нейронной сети, сформировав тестовой выборки.

В разделе два, также, описывается разработка прототипа программного обеспечения, которое осуществляется с использованием средств объектно ориентированного программирования C#, нейронная сеть была разработана в среде программирования MATLAB 2014, после чего была экспортирована в C# используя встроенные библиотеки MATLAB ComponentRuntime.

Третий раздел выпускной работы описывает процесс проведения вычислительных экспериментов разработанных методов и сравнение с существующими методами. Тестовые данные для проведения экспериментов были взяты в управлении сельского хозяйства муниципального района Волоконовский район. После сравнения разработанных методов, проводится сравнение метода BLUP и метода моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети. На

основе полученных результатов сравнения формируются рекомендации по улучшению и применению разработанных методов.

В заключении приведены основные результаты и выводы работы, подведены итоги результатов обучения и использования нейронных сетей, подтверждающие работоспособность разработанных методов и алгоритмов решения задач селекции.

Магистерская работа написана на 77 страницах, содержит 55 источников литературы.

# I Исследование современных подходов и информационных средств моделирования задач селекции

## 1.1 Обзор существующего информационного обеспечения моделирования задач селекции

Информационные технологии, уже привычное для общества слово, крепко вошедшее в нашу жизнь, стало одной из потребностей современного сообщества. Поэтому необходимо рассмотреть, что такое информационные технологии и информационное обеспечение и как они взаимодействуют с селекцией.

Информационные технологии — процессы, методы поиска, сбора, хранения, обработки, предоставления, распространения информации и способы осуществления таких процессов и методов [41].

Информационное обеспечение — совокупность показателей, справочных данных, классификаторов и кодификаторов информации, унифицированные системы документации, массивы информации, персонал, отвечающий за хранение, своевременность и качество технологии обработки информации.

Как видно из определений информационное обеспечение зависит от информационных технологий. Один из процессов связан с обработкой полученной информации, а другой процесс связан с получением качественно новой информации с использованием различных методов, характерные определенной сфере деятельности.

В информационных технологиях есть сложности для их применения для решения задач селекции. В области селекции — это отсутствие универсального метода, который безошибочно выдает результат по входным параметрам; многообразие селекционного отбора и их недостатки; сложности

технической реализации и многое другое. Все эти проблемы не дают возможность развиваться информационным технологиям в селекции.

Метод — это совокупность, действий, нацеленные на решение поставленной задачи и достижение иных целей.

В селекции растений особое значение имеет развитие методов и способов отбора растений, методов создания посевного материала, изучение мутагенеза, искусственного отбора и других видов наследования генов с целью выведения необходимого сорта растения. Все методы создаются под конкретные условия, который разрабатывается группой специалистов. Полученный метод решает совокупность задач, которые необходимы на момент разработки метода. Развитие науки подразумевает устаревание методов и необходимость их усовершенствования с использованием современных подходов и современных информационных технологий.

Перед решением поставленной задачи необходимо определить ее направление и объем. Направление характеризует конкретные вычисления или преобразования, а объем – количественные характеристики. Эти процессы связаны с получением необходимой информации для реализации проекта и хранения уже полученного результата.

Каждое информационное преобразование может быть осуществлено различными методами и средствами. На рисунке 1.1 показано разделение информационных технологий на методы.

Средства и методы информационных технологий		
Комплекс технических средств	Средства управления техническим комплексом	Организационно-методическое обеспечение

Рисунок 1.1 – Метод информационных технологий

Комплекс технических средств. Данный комплекс представляет собой совокупность инструментов и устройств, предназначенных для преобразования полученной информации.

Средства управления техническим персоналом. Комплекс средств позволяет персоналу наиболее эффективно использовать информационные технологии независимо от спектра решаемых задач для преобразования информации.

Организационно — методическое обеспечение связывает комплекс технических средств и средства управления в единый процесс в соответствии с требованиями задачи.

Так как в работе рассматриваются задачи селекции с использованием информационных технологий, необходимо выделить группы специалистов в данной области:

- 1) по обслуживанию технического оборудования;
- 2) программисты – создают новое информационное обеспечение;
- 3) специалисты по работе с готовым программным обеспечением.

Селекцию можно сравнить с процессом получения информации, ее преобразования и хранение конечного результата, что соответствует определению информационного обеспечения. Основываясь на этом, можно выделить несколько задач по преобразованию информации:

- 1) материалы по подготовке и оформлению документов в рамках решения задачи селекции;
- 2) нормативные документы по эксплуатации технических средств, документы по условию поддержания работоспособности оборудования;
- 3) инструктивные и методические материалы по организации работы персонала.

Совмещение информационных технологий в задачах селекции представляет совокупность человеческой деятельности, выполняемой

техническими средствами для осуществления потребностей общества в продукте.

Основное назначение информационных технологий:

- 1) повышение эффективности работоспособности;
- 2) расширение спектра трудовой деятельности;
- 3) исключение опасных для здоровья условий работы.

Для решения задач селекции наибольшую ценность представляют технические средства, позволяющие вести работу с использованием машин, которые частично или полностью автоматизируют решение задач селекции. Наиболее самостоятельным решением задач считается использование программ, выполняющие ряд поставленных задач.

С появлением земледелия, в качестве селекционера выступала эволюция. Так как данный процесс был очень длительным по времени, а потребности росли, то требовалась разработка новых методов, которые позволят сократить выполнение селекционной деятельности. Благодаря применению современного информационного обеспечения и технологий смоделировать процесс скрещивания родительских особей. Использование информационных технологий позволяет оптимизировать вычисления.

В селекционной деятельности обще доступными методами для использования является метод BLUP и программный модуль, который использует ИНС, по оценке насыщенности питательно смеси для растений.

Метод BLUP – метод селекционной и генетической оценки особей. Данный индекс определяется человеком и позволяет сделать вывод о практической ценности особи (продуктивность, генетика и д.р.).

BLUP — аббревиатура от английского («Best Linear Unbiased Prediction») лучший линейный неискаженный прогноз. На рисунке 1.2 представлена математическая модель на основе которой базируется выбранный метод.

Ключевые преимущества использования BLUP на практике:

- 1) разделение критериев, продуктивности:

- 2) влияние окружающей среды;
- 3) наследственность;
- 4) возможность одновременного сравнения параметров;
- 5) математический учет документально подтвержденных родственных связей;
- 6) корректировка всех значений племенной ценности по отношению друг к другу;
- 7) высокая точность племенной оценки;
- 8) возможность учета дополнительной информации.

$$Y_j = \mu + x_j^T \beta + \xi_j + \varepsilon_j,$$

$$\tilde{Y}_k = \mu + x_k^T \beta + \xi_k,$$

$$\hat{Y}_k = \sum_{j=1}^n c_{j,k} Y_j,$$

$$V = \text{Var}(\tilde{Y}_k - \hat{Y}_k),$$

$$E(\tilde{Y}_k - \hat{Y}_k) = 0.$$

Рисунок 1.2 – Математическая модель Best Linear Unbiased Prediction

Понятие BLUP было введено Чарльзом Роем Хендерсоном, профессором Корнельского университета, в то время речь шла о теоретической модели — абсолютно не приемлемой для практического использования ввиду крайней сложности вычислений. В конце 70-х начале 80-х годов, развитие науки и совершенствование компьютерных технологий, позволило быстро производить сложные математические расчеты.

Оценка племенной ценности необходима, чтобы корректно провести качество наследственной основы (например, высокий среднесуточный привес) в числовое выражение. Это обуславливает необходимость разработки соответствующих статистических методов, которые на основе собственной

продуктивности (фенотип) позволяют сделать заключение о генетической предрасположенности к определенной продуктивности.

Выбор наиболее лучших пород определяется на основе фенотипических значений по селекционному признаку, признак выбирается из целей селекции, с учетом их качественных показателей по интересующим параметрам. [20].

Как было сказано, методика BLUP является наиболее распространенной и доступной для исследования. Предварительно рассмотрим расчеты генетической ценности на основе метода BLUP. Результаты представлены в таблице 1.1 и характеризуют показатели выращивания скота.

Таблица 1.1 – Показатели фенотипических значений

Фенотипическое значение признака		Фенотипическое значение признака			
		I=2	I=3	I=4	I=5
Удой за год	P1	+0.34	+0.18	+0.08	-0.13
	P2	–	+0.38	+0.19	+0.13
	P3	–	–	+0.34	+0.31
	P4	–	–	–	+0.49
Содержание белка в молоке	P1	+0,37	+0,27	+0,18	+0,11
	P2	–	+0,33	+0,15	-0,07
	P3	–	–	+0,23	+0,02
	P4	–	–	–	+0,2

На основе данных таблицы 1.1, можно сказать, что для фенотипических показателей признаков (P<sub>i</sub>) наибольший уровень зависимостей отмечен в случае сопоставления показателей за смежные периоды. Однако при исследовании более удаленных по времени лактаций (например, 1-ой и 3-ей, 1-й и 4-ой, 1-й и 5-ой) происходило существенное снижение взаимосвязей всех показателей. В связи с этим, возникают сомнения в целесообразности отбора

выбранных особей. Таким образом отсеиваются особи, которые не подходят для разведения.

Второй метод для решения задач селекции основан на нейронных сетях, используется для оценки насыщенности питательной смеси. Нейронная сеть определяет достаточно точно и формирует таблицу, где имеется наименование и процентное содержание элементов. Полученные данные используются для корректирования питательной смеси.

На рисунке 1.3 приведена диаграмма концентрации отдельных питательных веществ в смеси с помощью нейронных сетей. Концентрация некоторых веществ велика, что требуется для улучшения усваиваемости.

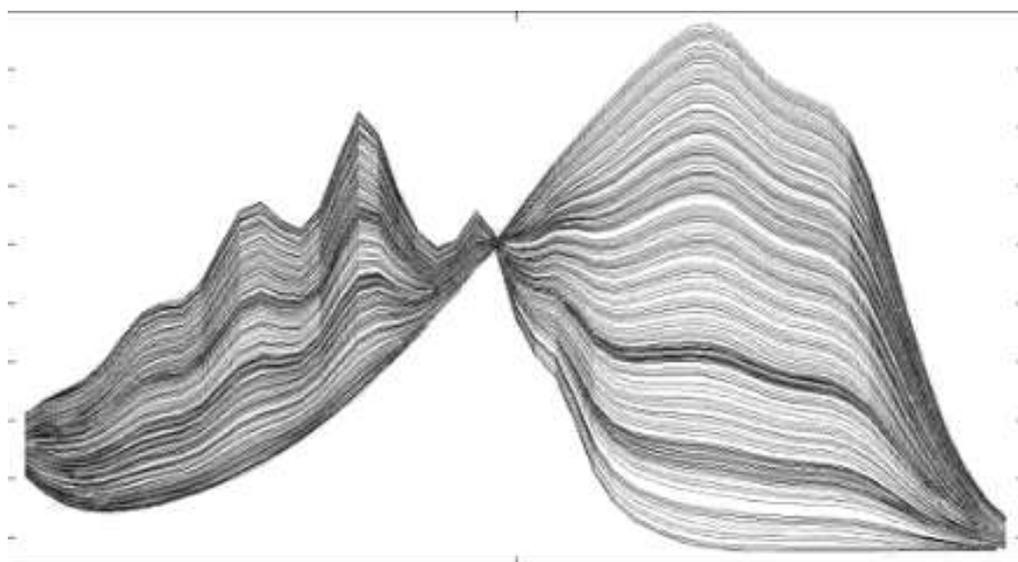


Рисунок 1.3 – Спектр концентрации элементов в корме

Поскольку этот метод не является самостоятельной программой, то предварительно необходимо обучать нейронную сеть, что является минусом использования метода. Плюсом данного метода можно выделить адаптивность под любые нужды.

Процесс обучения представленного метода характеризуется наличием тестовой выборки. В таблице 1.3 показана обучающая выборка для нейронной сети, где содержится перечень питательных элементов в кормовой смеси. В

качестве ожидаемого ответа, нейронная сеть должна дать спектр концентрации смеси (рисунок 1.3).

Таблица 1.2 – Обучающая выборка концентрации веществ

№ п/п	Смеси	
	Phe, мг/мл	Tyr, мг/мл
1	1,49	0
2	1,32	0,04
3	1,16	0,07
4	0,99	0,11
5	0,83	0,15
6	0,66	0,18
7	0,50	0,22
8	0,33	0,25
9	0,17	0,29
10	0	0,33

Данные из обучающей выборки предъявляются последовательно, вычисляется ошибка, затем веса подстраиваются для каждого набора данных, пока ошибка по всему обучающему массиву не достигнет низкого уровня.

## 1.2 Теоретические основы моделирования задач селекции

Первые зачатки селекции начали проявляться с появлением земледелия. Возделывая растения, человек выбирал среди собранного урожая лучшие и размножал их. Селекция проявлялась и в скрещивании древними людьми некоторых сортов винограда, пшеницы. Первым достижением селекции является сахарная свекла, которая содержит в три раза больше сахара

по сравнению с обычной свеклой. Достижения в ботанике помогают селекции развиваться.

В селекции, как науке, существуют несколько важных задач, которые она решает. Среди этих задач можно выделить выведение новых и совершенствование уже существующих методов.

Селекция – отрасль сельскохозяйственного производства, занимающаяся выведением новых сортов и гибридов культур, пород животных [1].

Основными направлениями селекции являются:

- 1) высокая плодовитость;
- 2) высокое качество плодов;
- 3) устойчивость к внешним раздражителям;
- 4) высокая скорость созревания;
- 5) низкие потребности в корме.

Для успешного решения представленных задач, академиком Вавилоновым Н. И, были выделены особые значения:

- 1) изучение сортового разнообразия интересующей для селекционера культуры;
- 2) влияние среды на развитие необходимых признаков;
- 3) изучение закономерностей наследования полученных признаков;
- 4) изучение изменчивости;
- 5) знание особенностей селекционного процесса;
- 6) грамотное применение стратегий искусственного отбора.

Достижения задач селекции позволило получить эффективные и экономически выгодные продукты, что позволило увеличить урожайность. В основном эти достижения связаны с сельскохозяйственными культурами, например, рис, пшеница. Они характеризуются высокой степенью выживаемости, высокой продуктивностью зерен.

Развитие селекции должно быть основано на законах генетики, поскольку генетика изучает наследственность и изменчивость. Не только

генетика важна в селекции, селекция использует накопленные знания и из других областей: эмбриология, биология, физиология.

Каждый новый сорт или вид организма должен учитывать потребности рынка. Так, например, имеет высокую популярность селекция насекомых и микроорганизмов, которые используются для борьбы с природными вредителями и возбудителями болезней культурных растений.

Под культурными растениями понимаются растения, которые выращиваются человеком. Как правило это продукты питания или текстильной промышленности. Такие растения не требовательны к ареалу и могут прорасти где угодно. Но не способны к само распространению как естественные растения или сорняки.

Говоря о потребностях рынка, мы имеем ввиду потребности рынка сбыта сельскохозяйственной продукции для удовлетворения имеющихся запросов потребителя. Селекция имеет целенаправленный характер. Например, пшеница способная давать урожай с высоким содержанием клейковины, для получения более качественной конечной продукции.

Успешное достижение поставленных целей во многом зависит от правильного выбора исходных материалов для селекции, особенностей для селекции, списка показателей, которые необходимо закрепить, выбранного метода скрещивания, изучения эволюционных факторов, учета требований организма ко внешней среде. С целью удовлетворения сортового разнообразия была создана коллекция культурных растений, которая хранится во Всесоюзном институте растениеводства и содержит более 400 тысяч образцов культурных и диких растений со всего земного шара.

Согласно данному справочнику, были установлены центры происхождения культурных растений (таблица 1.3).

Таблица 1.3 – Центры происхождения культур

Центр происхождения	Местоположение	Культивируемые растения
Южноазиатский	Тропическая индия,	Рис, сахарный
Восточноазиатский	Индокитай, острова	тростник, цитрусовые
Среднеземноморский	Юго–Восточной Азии	Соя, гречиха, слива,
Центральноамериканский	Китай, Япония, Корея,	вишня
Южноамериканский	Тайвань	Пшеница, рож, лен,
	Малая Азия, средняя	виноград
	Азия	Ячмень, кофе, бананы
	Южная Мексика	Кукуруза, какао, табак,
	Западное побережье	картофель.
	Южной Америки	

Наиболее разнообразными по количеству культур являются страны с древней историей, где наиболее длительно происходит процесс земледелия, отбор, гибридизация, полиплоидия, мутагенез, а также клеточная и генная инженерия.

В селекции существуют естественный и искусственный виды отбора. Искусственный отбор – процесс выбора человеком наиболее ценных сортов растений с целью их дальнейшего разведения или выведения нового сорта с желаемыми качествами. Данный отбор ведется с целью закрепления полезных для человека признаков у организма. Противоположностью искусственному отбору служит естественный отбор, который повышает приспособляемость организмов к новым условиям. Сравнение существующих видов отбора сортов растений представлена в таблице 1.4.

Таблица 1.4 – Сравнительная характеристика видов отбора

Признак	Естественный отбор	Искусственный отбор
Исходный материал для отбора	Индивидуальные признаки организмов	Индивидуальные признаки организмов
Отбирающий фактор	Условия среды (живая и неживая природа)	Человек
Путь благоприятных изменений	Остаются, накапливаются, передаются по наследству	Отбираются, становятся производительными
Путь неблагоприятных изменений	Уничтожаются в борьбе за существование	Отбираются, бракуются, уничтожаются
Направленность действия	Отбор полезных признаков	Отбор признаков, полезных человеку
Результат отбора	Новые виды	Новые сорта растений, породы животных
Формы отбора	Движущий, стабилизирующий	Массовый, индивидуальный, бессознательный

Достижение высоких целей в селекции невозможно без информационного обеспечения, которое содержит способы и методы развития и изучения отраслей науки.

К классическим методам селекции растения относятся гибридизация и отбор. Различают несколько форм отбора:

Массовый отбор. Отбор, применяемый при перекрестно опыляемых растений. Например, рожь, кукурузу, подсолнечник. При этом идет выделение наиболее ценную группу. В этом сорте растений у гетерозиготных особей каждое себя будет обладать уникальным генотипом. Массовый отбор способствует улучшению и сохранению качеств сорта, но полученные результаты отбора неустойчивы.

Индивидуальный отбор. Этот вид отбора применим для само опыляемых растений. В этом случае полученное потомство сохранит признаки родителя, а плоды будут гетерозиготные.

Инбридинг. Данный метод используется при самоопылении перекрестно опыляемых растений. Является гибридным методом первых двух. Необходим для получения чистых линий, например, кукурузы. При этом подбираются гибриды с максимальной жизненной силой. Закрепляют полученный эффект путем скрещивания чистых линий между собой.

На рисунке 1.4 показана кукуруза, которая подверглась инбридингу. Центральный початок кукурузы с наибольшей жизненной силой. Слева и справа початки с чистой линией родительской формы.

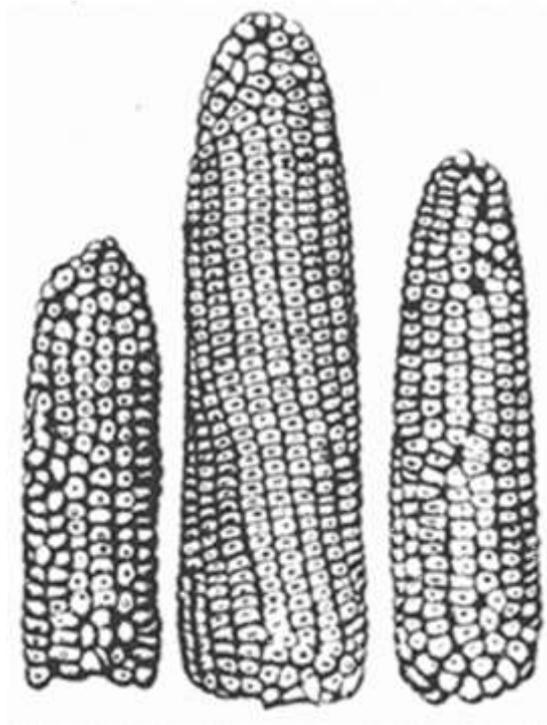


Рисунок 1.4 – Виды кукурузы

Наличие высокого уровня жизненной силы у плода объясняется как гипотезой доминирования, когда чем больше в генотипе генов, тем больше эффект гетерозиса и большой коэффициент урожайности.

$$P \text{ AA}bbCCdd \times aaBBccDD \text{ F}_1 \text{ AaBbCcDd}$$

Гипотеза сверхдоминирования, когда гетерозиготное состояние имеет преимущество на родителей по продуктивности. У эффекта гетерозиса есть недостаток, начиная со второго поколения он начинает затухать, а гены переходят в гомозиготное состояние.

$$Aa \times Aa$$

$$AA \ 2Aa \ aa$$

На рисунке 1.5 показан росток гречихи. Слева образец с диплоидным набором генов ( $2n=16$ ), а справа с тетраплоидным набором генов ( $2n=32$ ).



Рисунок 1.5 – Гречиха

Перекрестное опыление самоопылителей. Отбор дает возможность к сочетанию различных сортов. Рассмотрим на примере создания новых сортов пшеницы. На цветке растения удаляются пыльники, и рядом с ним ставят другой сорт и герметично накрывают сосудом. В результате получается новый сорт с гибридными семенами, сочетающие признаки обоих сортов пшеницы.

Отдаленная гибридизация. Межвидовое скрещивание растений. Отбор, в результате которого получают стерильные гибриды с нарушением мейоза, и у них не образуются гаметы. Примером такого отбора является тритикале, показанные на рисунке 1.6.

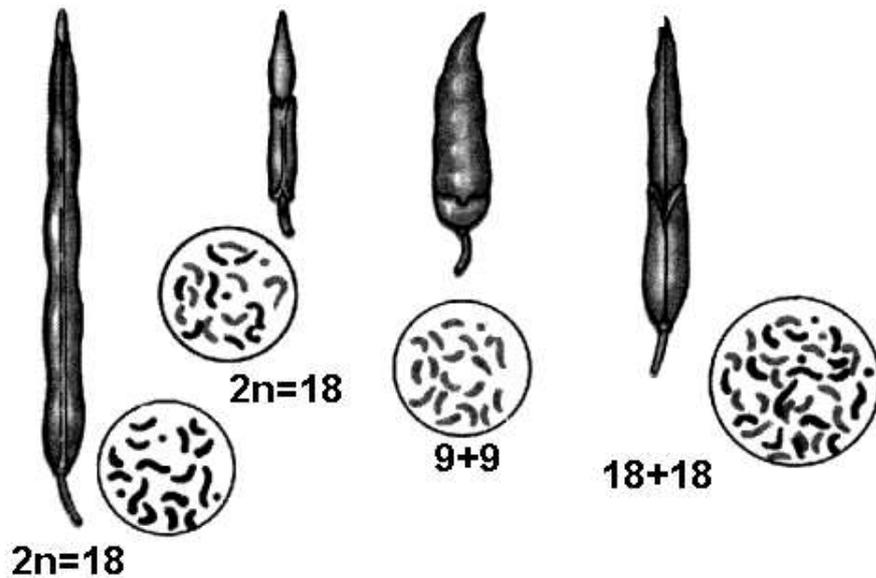


Рисунок 1.6 – Тритикале

На рисунке 1.6 показан пример отдаленной гибридизации. Хотя такие сорта и стерильны, но их плодовитость можно восстановить с использованием колхицина. С его помощью удалось удвоить количество хромосом каждого представителя в гибриде, тем самым восстановив плодовитость.

Многие методы были предложены Мичуриным И. В. С помощью рассмотренных методов решались проблемы человечества с обеспечением питания. Все представленные методы имеют проблемы, которые со временем решаются.

Современные технологии позволяют воплотить представленные методы селекции. Для этого необходима специальная программа и полный перечень входных параметров. Информационные технологии позволяют сохранять, обрабатывать и моделировать результаты опытов в селекции и выдавать рекомендации по улучшению полученного результата.

### 1.3 Обзор структур и методов обучения нейронных сетей

Применяемые технические средства в задачах селекции способствуют получению новой информации, тестированию возможностей и т.д. Существующие программные комплексы либо тяжелы в настройке, либо имеют не первостепенное значение в получении нового вида растения. Использование нейронных сетей позволит гибко использовать такой инструмент под различные условия проведения тестов или выведения условий существования сорта.

Императивное программирование – это вид создания программного обеспечения, при котором программе представляется проблема и путь ее решения. В случае появления ошибок, программа не способна подстроиться для их предотвращения или анализировать полученный результат без вмешательства разработчика.

Написанная таким образом программа характеризуется следующими параметрами:

- 1) в коде программы записываются поэтапные команды;
- 2) инструкции выполняются последовательно;
- 3) при повторном выполнении инструкции могут считываться из памяти;
- 4) данные могут быть записаны в память.

Использование данного подхода нагружает код и увеличивает возможность появления ошибок [4]. В связи с этим недостатком, программное обеспечение, основанное на императивной парадигме программирования, оказалось не эффективным при анализе входящих данных из-за загруженности программного кода. Поэтому началась разработка новой парадигмы программирования.

Декларативное программирование – это стиль программирования, при котором описывается проблема и ожидаемый результат. Разработанная программа самостоятельно выполняет все вычисления и манипуляции с

данными, а полученный результат выводит пользователю. Программное обеспечение, основанное на использовании такой парадигмы программирования, наиболее гибко настраивается и способна обучаться [19].

Новая парадигма программирования помогла создавать программы, которые отдаленно напоминают человеческое мышление, а конечный код не так громоздким. Программные продукты были способны на анализ, группировку и многое другое, что показало, насколько эффективна новая парадигма. Созданная парадигма легла в основу искусственного интеллекта, экспертных систем и искусственных нейронных сетей.

Каждая программа, как и метод, создается под определенные нужды и для определенного слоя или рода деятельности человека. На рисунке 1.7 показано разделение программного обеспечения, предназначенного для решения задач пользователя – прикладное ПО.



Рисунок 1.7 – Прикладное программное обеспечение

Прикладное программное обеспечение разделяется на:

Программное обеспечение общего назначения – совокупность программ для решения общих универсальных задач. Эти программы используются большинством пользователей компьютера;

Программное обеспечение специального назначения – совокупность программ для решения более узких задач и профессиональных задач различных предметных областей.

Программное обеспечение для решения задач селекции относится ко второму типу, так как решает задачи профессионального характера предметной области. К подобным программным продуктам, относятся продукты с экспертными системами.

Искусственная нейронная сеть (ИНС) — математическая модель, построенная по принципу организации и функционирования биологических нейронных сетей нервных клеток живого организма [23].

Экспертная система — компьютерная система, которая в состоянии частично или полностью заменить человека при решении поставленной задачи. Экспертные системы разрабатываются исследователями искусственного интеллекта. В 1980-х годах получили коммерческое подкрепление [15].

К экспертным системам, также, относятся и искусственные нейронные сети, которые способны анализировать ситуацию и выдавать решение в зависимости от направленности задачи. Отличием нейронных сетей от экспертной системы является их универсальность при решении различных задач. Дополнительно к нейронной сети необходимо создать базу знаний, на которой экспертная система будет обучаться анализировать и делать выводы.

Рассмотри структуру интеллектуальных систем:

- 1) интерфейс;
- 2) база знаний;
- 3) инженер базы знаний;
- 4) механизм вывода;
- 5) обучающая подсистема.

На рисунке 1.8 представлено биологическое представление нейрона и его связи с другими нейронами и формирование нейронной сети живых организмов.

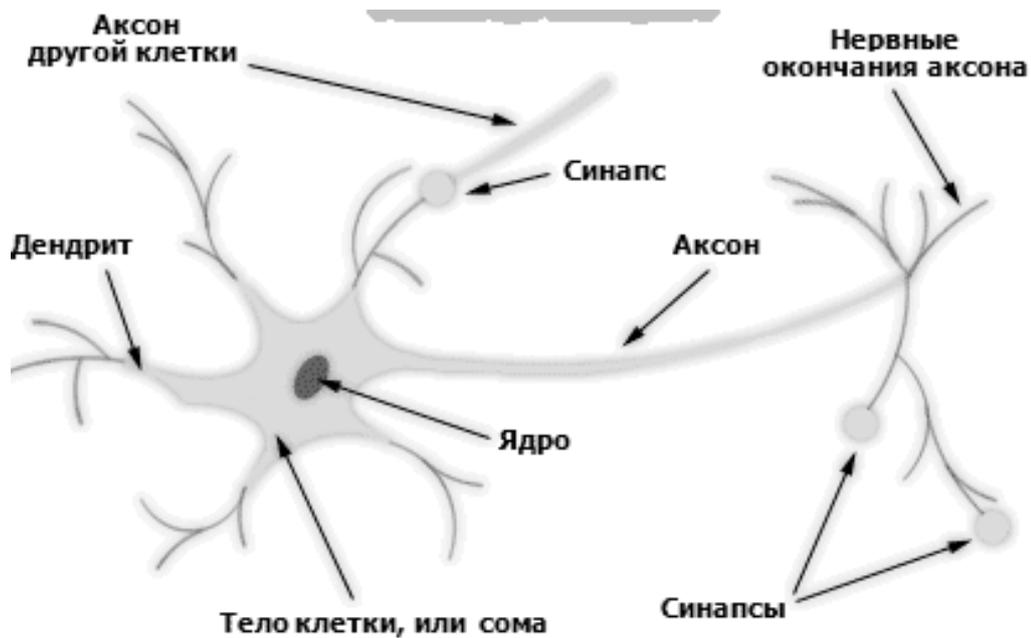


Рисунок 1.8 – Упрощенная модель биологической нейронной сети

Разработка нейронных сетей опирается на природные аналоги (структура человеческого мозга), согласно ему, в любая нейронная сеть должна обладать персептроном – основной элемент с  $N$  входами и  $M$  выходов. Взаимосвязь входа и выхода определяется целями и задачами область, в которой применяется нейронная сеть. Обычно эта связь определяется по формуле 1.1:

$$y(n) = \varphi\left(\sum_{i=1}^N c_i x_i(n)\right) \quad (1.1)$$

где  $c_i$  – параметр персептрона в момент подачи  $x_i$  сигнала;

$n$  – количество персептронов;

$x_i$  – сходной сигнал.

Комбинации созданных персептронов создают нейронную сеть, а их связи образуют структуру. Наибольшую популярность получили статические и динамические многослойные нейронные сети. Основной характеристикой является наличие нескольких слоев, взаимодействующие друг с другом с помощью передаваемого сигнала от внешнего слоя с меньшим номером ко

внутреннему слою с большим номером. На рисунке 1.9 наглядно показан описанный процесс.

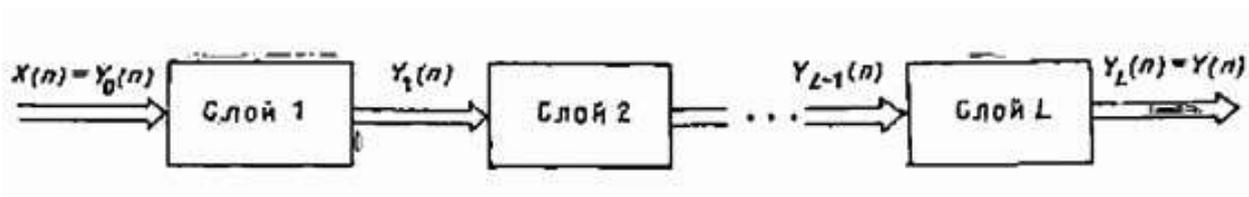


Рисунок 1.9 – Структура многослойной нейронной сети

Рисунок 1.9 демонстрирует статическую сеть. На рисунке 1.10 показана структура динамической многослойной сети.

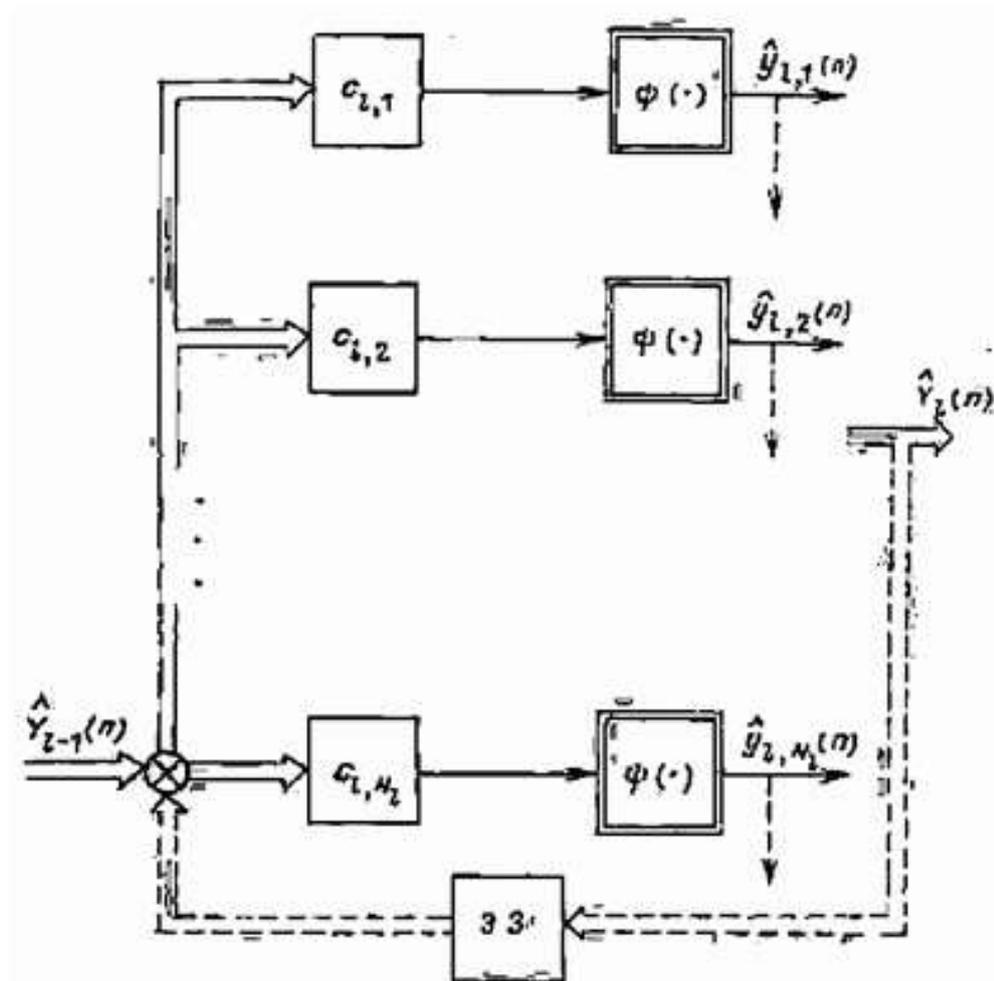


Рисунок 1.10 – Структура нейронной сети

Их различие заключается в том, что статические НС служат для аппроксимации статических объектов, а динамические — для аппроксимации или построения моделей нелинейных динамических объектов [45].

Независимо от использованной структуры НС, ее необходимо обучать, тестировать и только потом проводить апробацию созданной структуры на реальной выборке данных. В процессе обучения нейронная сеть «настраивает» веса персептронов посредством моделирования среды, в которую эта сеть встроена. На рисунке 1.11 показано строение персептрона.

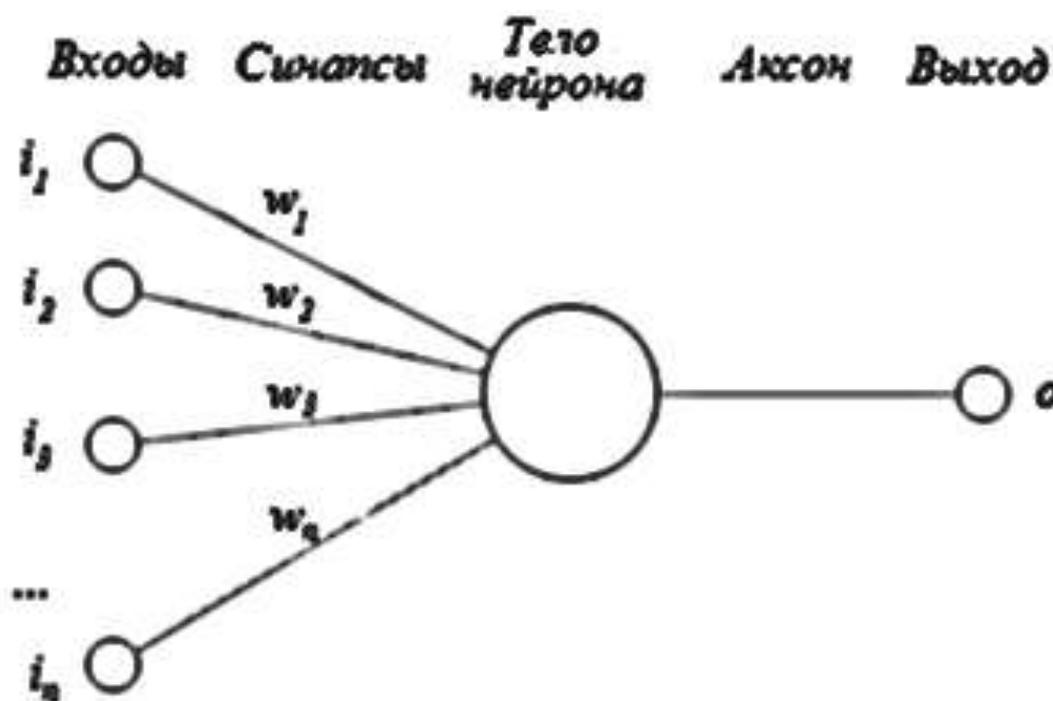


Рисунок 1.11 – Схема персептрона

Основными элементами нейронных сетей являются:

- 1) входы;
- 2) синапсы;
- 3) тело нейрона;
- 4) аксон;
- 5) выход.

При соблюдении всех этапов построения нейронной сети, она будет способна обучаться, что повысит ее эффективность.

Необходимо выбрать метод обучения нейронной сети. Тип обучения определяется способом подстройки параметров. Существует два типа обучения: обучение с учителем и без учителя.

При обучении без учителя обучающая выборка состоит лишь из входных данных. Алгоритм нейронной сети подстраивает веса сети таким образом, чтобы получались согласованные данные. Другими словами, процесс обучения выделяет статистические свойства обучающей выборки и группирует сходные данные в классы для дальнейшего распознавания классификации реальных данных в реальных условиях. На рисунке 1.12 приведена многослойная нейронная сеть.

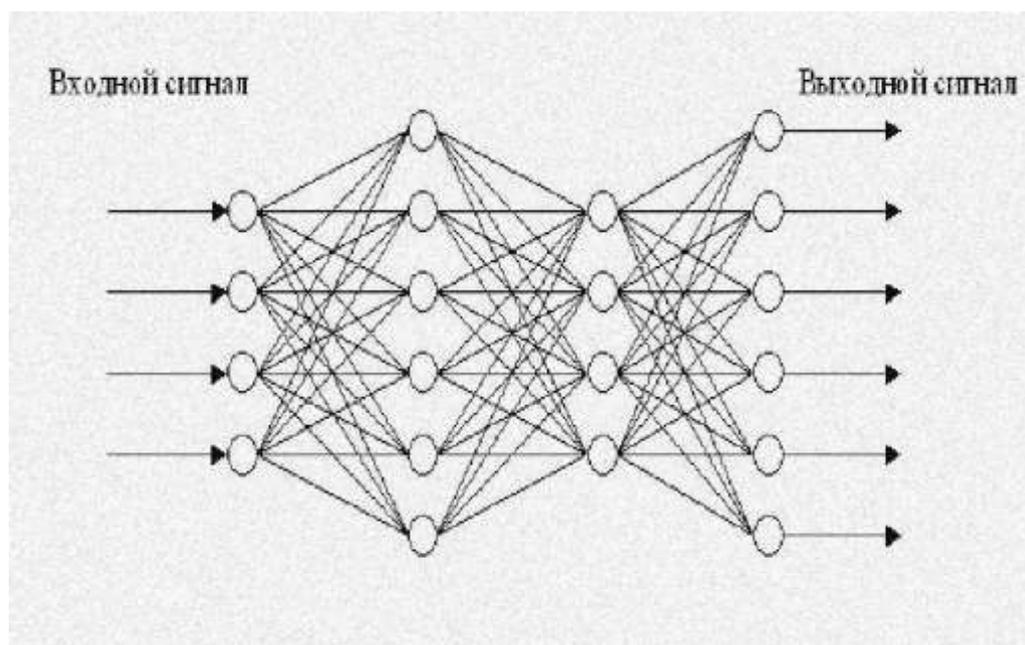


Рисунок 1.12 – Многослойная сеть

Для разработки методов скрещивания сортов растений необходимо использовать нейронную сеть. Дополнительной задачей является создание методов эффективного скрещивания на основе изучения реальных методов, применимых на производстве. Совершенствование приемов селекции требует учета множества факторов.

Для достижения эффективного использования нейронных сетей, целесообразно использовать многослойную нейронную сеть, что позволит расширить область применения и повысить качество обучения. Многослойная сеть значительно повышает точность и скорость решения, но увеличит количество времени необходимое для обучения такой сети.

Многослойная нейронная сеть содержит в себе входной/выходной слои и несколько вычислительных слоев, которые соединены с одним выходным нейроном. Для эффективного использования возможностей нейронных сетей целесообразно использовать особенности их функционирования.

Первым важным свойством нейронных сетей является высокая скорость их работы. Примером скорости работы нейронных сетей является их использование при работе автопилота автомобиля. Нейронная сеть распознает окружающие объекты достаточно быстро и на выходной слой передает команды дальнейших действий. Примером точности решения является использование нейронных сетей в задачах определения рисков в информационной среде.

Обучение нейронной сети предполагает наличие метода и процесса ее обучения. Подбор процесса обучения является одной из задач улучшения работы нейронной сети. Наиболее распространен следующий тип обучения – метод обратного распространения ошибки. Указанный подход к обучению характеризуется следующими действиями:

- 1) корректировка веса для минимизации ошибки;
- 2) сравнение с заранее подготовленной базой образов;
- 3) обучающая выборка.

Обучение нейронной сети требует значительного времени. Необходимым условием для успешного обучения НС является наличие двух массивов данных. Первый массив данных – это обучающие выборки (учебные примеры) для системы, второй – используется для оценки качества работы нейронной сети и последующей оценке ошибки ее работы. Обучение позволяет сравнивать полученный результат с желаемым и уточнять значения

весов нейронов. На рисунке 1.13 схематично описан процесс обучения нейронной сети.

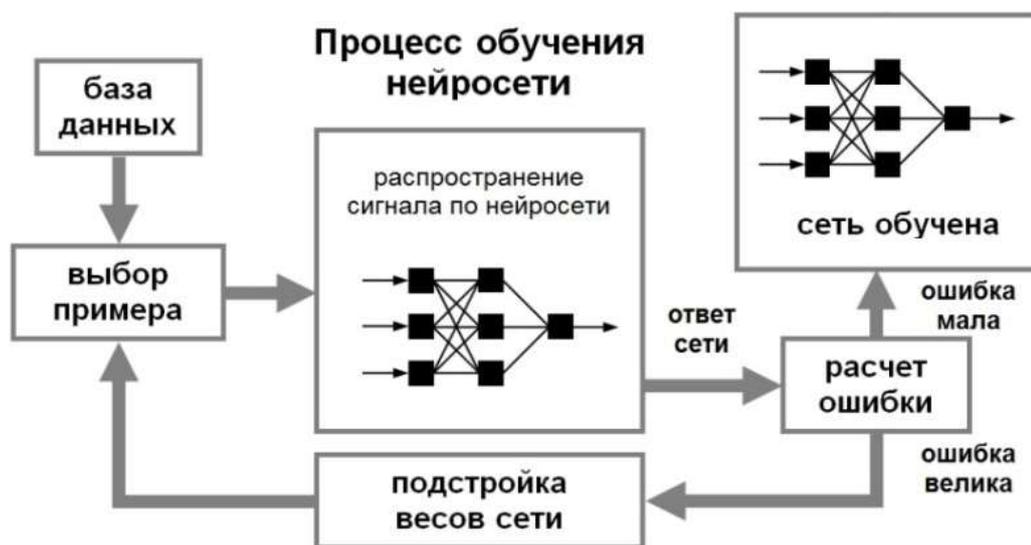


Рисунок 1.13 – Процесс обучения нейронной сети

Указанные на рисунке этапы обеспечивают выполнение следующих задач:

- 1) инициализация синоптических весов;
- 2) корректировка веса сети;
- 3) вывод информации.

При проведении исследований в области селекции авторы использовали нейронную сеть для определения концентрации элементов в кормах, которая обучалась с использованием метода обратного распространения ошибки. Также нейронные сети планируется применять для определения степени испорченности кормов и контроля за проведением скрещивания и получения желаемого результата.

В работе уже не раз упоминалось обучений нейронной сети, рассмотрим процессы обучения нейронной сети. Способность к обучению является особенностью человеческого мозга. Поскольку нейронные сети разрабатывались с опорой на биологический мир, то и их является возможным

обучать. В контексте искусственных нейронных сетей процесс ее обучения рассматривается как настройка используемой архитектуры сети, весов связей для наиболее эффективного выполнения поставленной задачи. Настройка весов происходит по предоставленным обучающим выборкам.

Среди существующих методов обучения выделяют: детерминированный и стохастический методы обучения.

**Детерминированный метод.** Данный метод обучения итеративно корректирует параметры сети, основываясь на ее текущих параметрах, количестве входных нейронов, и выходных нейронов. Ярким примером данного метода является метод обратного распространения ошибки.

**Стохастический метод.** Этот метод обучения изменяет параметры сети случайным образом, сохраняя изменения, которые привели к улучшению результата. Алгоритм стохастического метода:

Шаг 1. Выбор значений для параметров случайным образом. Корректировка их значений. Предъявить множество входов и вычислить получающиеся выходы.

Шаг 2. Сравнить получившиеся выходы с желаемыми и вычислить разницу между ними. Полученное значение является ошибкой получения выходного результата.

Шаг 3. Если ошибка уменьшилась, сохраняем значения параметров, иначе коррекция останавливается и начинается новая.

Шаги 2 и 3 повторяются пока сеть не обучится.

У данного метода есть недостаток, который связан с проблемой локальных минимумов (рисунок 1.16), когда результат может попасть в «ловушку» локального минимума.

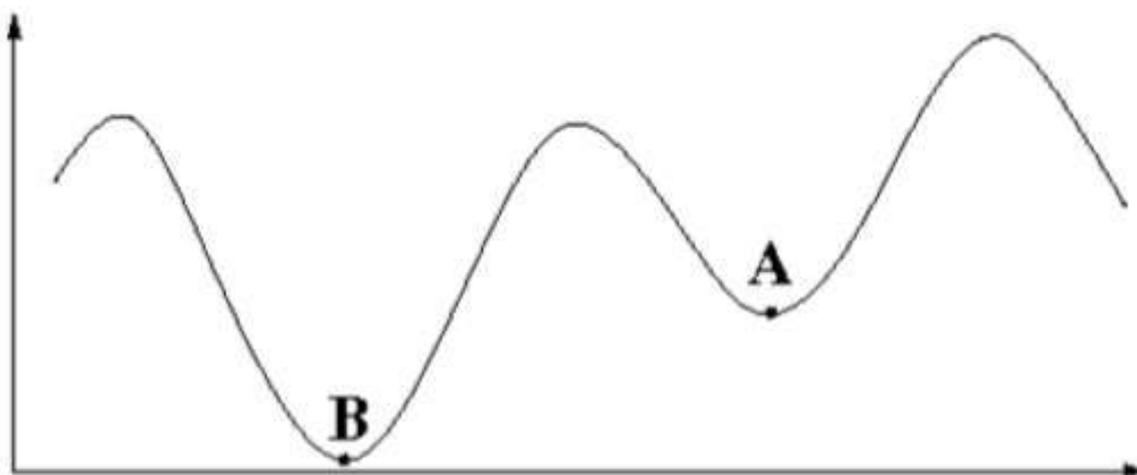


Рисунок 1.14 – Проблема локальных минимумов

Что бы избежать подобной ситуации, можно постепенно уменьшать средний размер случайных шагов коррекции. Если плавно уменьшить размер шагов, то будет достигнуто условие, при котором значение ошибки будет на некоторое время застывать в точке «В». Когда размер шага уменьшится еще больше, значение ошибки будет останавливаться на короткое время в точке «А», и в точке «В». Если уменьшать размер шага непрерывно, то будет достигнута величина шага, для преодоления локального минимума «А».

Рассмотрим процесс обучения с учителем. Для того чтобы алгоритм являлся таковым, необходимо наличие правильных значений на каждый входной пример. В процессе обучения сеть меняет свои параметры, чтобы получить нужный коэффициент на выходе. При этом, размер множества должен быть достаточным для обучения нейронной сети [46].

Метод обучения без учителя. Из-за того, что метод обучения с учителем подвергается большой критике, появился метод без учителя, который может применяться тогда, когда известны только входные сигналы. На основе входных сигналов сеть обучается выдавать наиболее подходящее значение выходных нейронов. Обычно алгоритм подстраивает значения таким образом, чтобы сеть выдавала одинаковые результаты для достаточно близких выходных значений [40].

Метод Хэбба. Суть данного метода заключается в предположении Дональда Хэбба, что вес соединения между двумя нейронами усиливается, если оба нейрона находятся в возбужденном состоянии.

Алгоритм выглядит следующим образом:

Шаг 1. На стадии инициализации нейронной сети всем весовым коэффициентам присваиваются небольшие случайные значения.

Шаг 2. На вход подается входной сигнал и вычисляется выход.

Шаг 3. На основании выходного значения производится изменение весовых коэффициентов.

Шаг 4. Повтор шагов 2 с новым входным параметром пока выходные значения сети не синхронизируются с заданной точностью.

Обучение по правилу коррекции по ошибке. Данный метод обучения разработан в 1957 году. Метод базируется на методе обучения с учителем, только в этом методе для каждого входного примера задается желаемый выход. Если входное значение не совпадает с желаемым, то весовые коэффициенты и другие параметры сети будут скорректированы. Коррекция производится на разницу между реальным и желаемым выходом сети.

Генетические алгоритмы. К генетическим алгоритмам относятся алгоритмы, основанные на моделировании развития биологической популяции.

В таком алгоритме популяцией называется набор векторов  $P = \{p^i\} = \{p^1, \dots, p^N\}$ , где  $N$  – размер популяции. Элементы  $p^i$  – особи популяции. Каждый вектор  $p^i$  содержит все параметры, с помощью которых можно описать каждую особь.

Существует функция  $E(p)$ , зависящая от вектора  $p$ , с помощью которой вычисляется ошибка. Требуется найти минимум  $E$ . Элементы множества  $P$  способны развиваться согласно правилам:

Если  $E(p^0)$  – мало, то особь  $p^0$  является удачной и получает высокий приоритет дальнейшего размножения. Вероятность гибели уменьшается.

Если  $E(p^0)$  – велико, то особь  $p^0$  является неудачной и получает низкий приоритет дальнейшего размножения. Вероятность гибели увеличивается.

В результате скрещивания любая особь популяции может мутировать, т.е. происходит смещение ее значения на величину  $p^0 := p^0 + \Delta p$ , где  $\Delta p$  – небольшой по модулю вектор, характеризующий мутацию особи.

На стадии размножения каждая точка делится. Правила такого деления определяется модель размножение или селекции. В случае гибели особи, она удаляется из множества  $P$ . Определение коэффициентов происходит экспериментальным путем.

#### 1.4 Анализ эффективности применения нейронных сетей

Анализ нейронных сетей показал, что они способны к обучению, а, следовательно, возможно их применение практически в любой сфере деятельности, что является положительной стороной нейронных сетей. Разнообразие областей применения ИНС определяет их эффективность в решении задачи. Среди областей их применения можно выделить:

Распознавание образов. Нейронная сети способна анализировать различные объекты, при обучении, нейронной сети указывается объект, и группа к которой он относится. Показанный образец является базой знаний для ИНС. По окончании обучения сети ей можно предъявлять неизвестные ранее образы и получать ответ о принадлежности к определённому классу.

Принятие решений и управление. Данная область применения близка к задаче классификации, где нейронная сеть классифицирует объект на основе его характеристик.

Прогнозирование. Прогнозирование – это способность нейронной сети к обобщению и выделению связей. По завершению обучения сеть способна предсказать будущее значение некой последовательности на основе

нескольких предыдущих значений и (или) каких-то существующих в настоящий момент факторов.

В результате анализа областей применения ИНС мы установили, что нейронные сети способны эффективно работать практически в любой области, но в сочетании с другими методами и правилами:

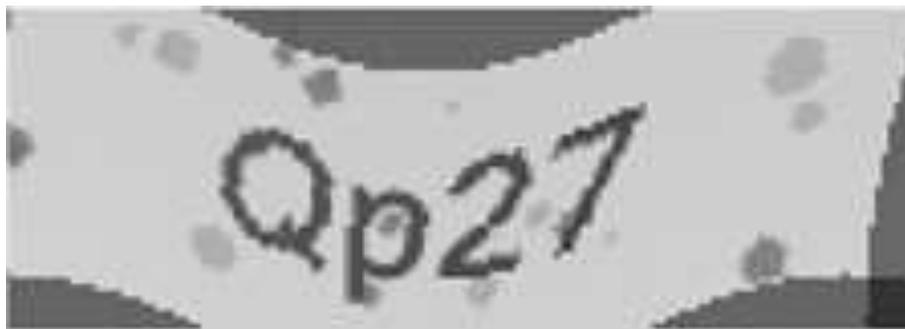
- 1) методы индивидуального отбора растений;
- 2) методы создания персептрона;
- 3) методы анализа входных параметров

Также положительной стороной нейронных сетей является возможность их создания практически на любом языке программирования. Достаточно лишь описать структуру нейронной сети, механизм обучения и ее тип: с учителем или без учителя. Не зависимо от этого необходимо ее первичного обучения.

База знаний создается следующими группами людей:

- 1) эксперты в проблемной области;
- 2) специалисты по разработке ИНС;
- 3) программисты, осуществляющие реализацию ЭС.

Приведем пример внедрения нейронных сетей для распознавания объектов. На рисунке 1.14 показан результат распознавания нейронной сети.



Qp27

Рисунок 1.15 – Объекты распознавания

Распознавание происходит не всегда с первого раза, а некоторые объекты могут быть не распознаны совсем. Для улучшения результатов, нейронную сеть оптимизируют и улучшают обучающую выборку. После всех манипуляций получаем следующие результаты [12] (рисунок 1.15).

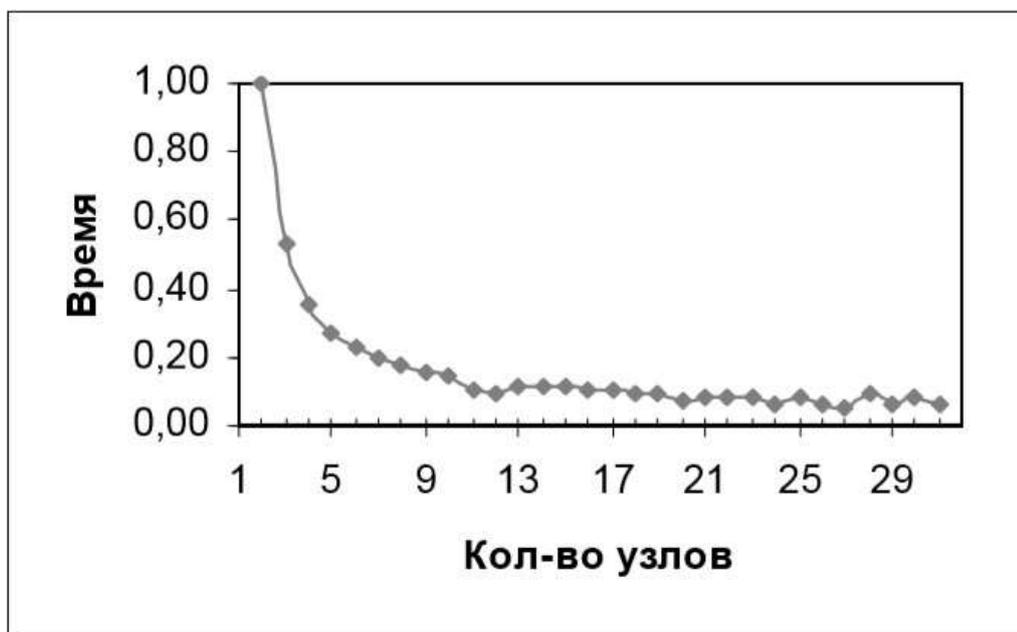


Рисунок 1.16 – Улучшение эффективности работы нейронной сети

На рисунке 1.15 показано изменение времени решения задачи с увеличением вычислительных узлов.

Исходя из многообразия областей, где применяются искусственные нейронные сети можно утверждать о высокой эффективности их применения для решения задач, а, следовательно, целесообразно их применения для решения задач селекции. Где их применение поможет анализировать входные данные, выдавать рекомендации, находить взаимосвязи и создавать новые методы селекции растений с меньшими временными затратами для получения готового результата.

На основании рассмотренного теоретического материала по исследованию информационного обеспечения зада селекции, а также методов селекции и анализа существующих программных решений было установлено следующее. Развитие информационных технологий находит на высоком уровне, так же, как и теоретическое описание методов и видов селекции. На недостаточном уровне разработанности находится практическая реализация разработанных методов решения задач селекции. Анализ существующих алгоритмов нейронных сетей, методов их обучения и областей применения, показал, что развитие нейронных сетей имеет большой потенциал и широкую область применения. Описание областей применения нейронных сетей привело к выводу, что для каждой задачи существует нейронная сеть. А анализ способов обучения, позволил выбрать генетический алгоритм как основу для обучения нейронной сети.

## 2 Разработка методов и алгоритмов моделирования процесса селекции

### 2.1 Разработка методов моделирования селекции

В селекции растений доступны только методы, представленные в первом разделе. Анализ опытов работы с существующими методами, показал необходимость наличия параметров для работы метода, по которым будет происходить формирование гена для манипуляций с ним. Параметры должны максимально характеризовать растения. Данные для каждой особи заполняются индивидуально. Для исключения ошибок на этапе заполнения контрольных значений, введем таблицу, которая будет служить шаблоном. В таблице 2.1 будут храниться начальные данные для работы методов.

Таблица 2.1 – Показатели фенотипа родителей

Родители Параметр	1	2	...	n
1	$C_1$ 1	$C$ 12	...	$C$ 1n
2	$C_2$ 1	$C$ 22	...	$C$ 2n
...	...	...	...	...
m	$C_m$ 1	$C$ m2	...	$C$ mn

В этой же таблице необходимо добавить ключевые параметры, по которым будет происходить ранжирование всего списка. Особенность созданной структуры входных данных является возможность проследить формирование гена по мере ввода данных.

Разделение родительских особей будет происходить на основе методов, представленных ниже.

Рулетка. Метод отбора особей основанный на колесе рулетки. При этом родители выбираются исходя из значений приспособленности организма.

Турнир. Данный метод подразумевает разделение всего списка родителей на небольшие группы по 2-3 особи в группе. После чего из группы выбирается особь случайным образом. Пример работы показан на рисунке 2.1.

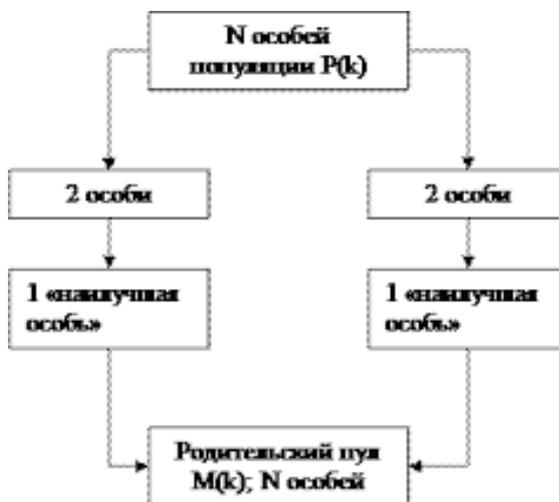


Рисунок 2.1 – Турнирная селекция

Представленная схема демонстрирует работу методов для групп, которые состоят из двух особей. Схема работы методов легко применима и к группам большего размера.

Ранговая. Каждой особи в группе присваивается значение, характеризующее уровень приспособляемости, на рисунке 2.2 показан процесс сортировке. Остается только выбрать особей с необходимым показателем.

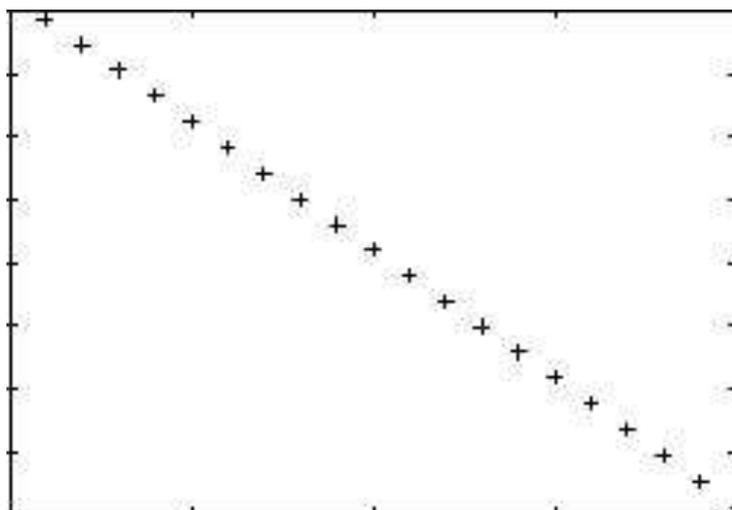


Рисунок 2.2 – Ранжирование особей

Исходные данные будут отсортированы по ключевому признаку, поэтому механизмом разделения выбрана ранговая система.

На этапе выбора структуры нейронной сети и метода ее обучения, был выбран генетический алгоритм для обучения, которому необходимо условие для работы. Таким условием является достижение от каждой пары не более четырех новых особей.

Переходим к процессу скрещивания. Скрещивание является важным этапом селекции, именно при скрещивании осуществляется обмен генетической информацией между двумя особями. Потомки содержат в себе комбинацию генетических особенностей обоих родителей. Эффективность работы оператора скрещивания зависит от эффективности операции кроссинговера, как и успешность генетического алгоритма. В рамках решаемой задачи реализован оператор кроссинговера, который осуществляет поэтапное преобразование генетического кода особи.

Генетический код особи кодируется двоичным кодом или кодом Грея. Стоит отметить, при двоичной кодировке параметров теряется точность параметров, что приведет к неверному результату декодирования. Недостаток заключается в том, что соседние числа отличаются в значениях нескольких битов. Решением данного недостатка является использование кода Грея, реализация этого кода представлена в таблице 2.2.

Таблица 2.2 – Кодировка Грея

Десятичный код	Двоичное кодирование	Кодирование по коду Грея
	двоичное значение	двоичное значение
1	2	3
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100

Продолжение таблицы 2.2.

1	2	3
9	1001	1101
10	1010	1111
11	1011	1110
12	1100	1010
13	1101	1011
14	1110	1001
15	1111	1000

Таким образом, при кодировании целочисленного признака мы разбиваем его на тетрады и преобразуем по коду Грея.

Рассмотрим алгоритм кодировки Грея:

- 1) разделить интервал значений признака на участки;
- 2) представить значение каждого гена в целочисленном виде (используя код Грея);
- 3) принять за значение параметра середину интервала.

В качестве примера приведем признак, значение которого находится в интервале от 0 до 1. При кодировании участок разделяется на 256 интервалов, число можно выбрать иное. Для кодирования их номера таким образом потребуется 8 бит. Допустим значение гена: 00100101bG (заглавная буква G показывает, что используется кодирование по коду Грея). Для начала,

используя код Грея, найдем соответствующий ему номер интервала: 25hG/36h/54d. Теперь посмотрим, какой интервал ему соответствует. После несложных подсчетов получаем интервал  $[0,20703125, 0,2109375]$ . Значит значение нашего параметра будет  $(0,20703125+0,2109375)/2=0,208984375$ .

Иначе говоря, чтобы определить все параметры объекта, необходимо знать значения генов. Выполнив все преобразования, получим закодированную таблицу начальных данных, представленную в таблице 2.3.

Таблица 2.3 – Значения признаков

Родители Параметр	1	2	...	n
Параметр 1	0010	0011	...	0100
Параметр 2	1010	1100	...	1101
...	...	...	...	...
Параметр m	1001	0111	...	1110

Таким образом, мы перевели значения параметров родителей согласно алгоритму кодировки Грея.

Процесс скрещивания родительских генов показан на рисунке 2.3, где две родительские цепочки хромосом образуют новых потомков. Далее определяется место разрыва цепочки, после чего левая часть «передается» новому поколению. Правая часть нового поколения получается за счет такой операции, но над второй родительской генетической цепочкой.

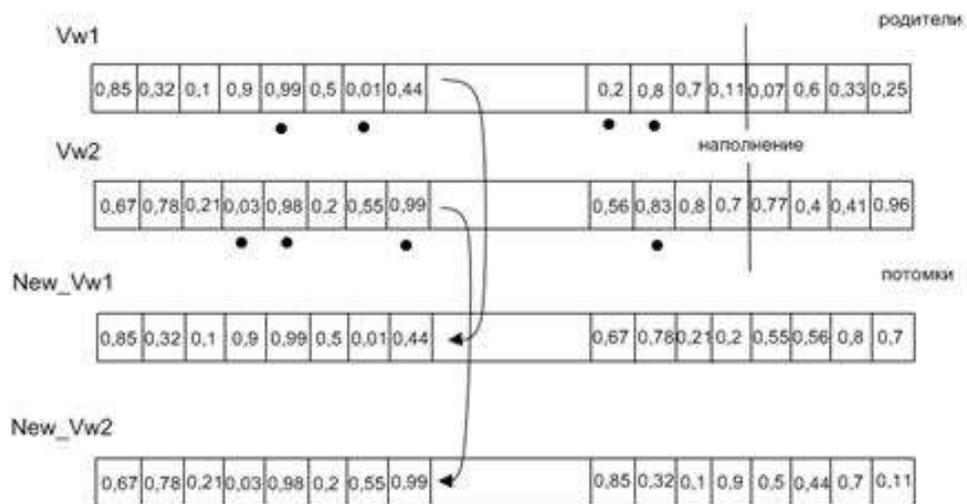


Рисунок 2.3 – Кроссинговер особей

Аналогичная процедура выполняется при получении второго потомка New\_Vw2, который наследует левую часть второй родительской хромосомы. Заполнение оставшихся генов второго потомка осуществляется за счет информации, находящейся у второго родителя. Таким образом из двух родительских генов получается две новые особи.

Процедура обмена генов основывается на следующем соотношении (формула 2.1):

$$\begin{cases} x_i = \text{random}(1), i = \overline{1, l}; \\ n = \arg \max x_i, i = \overline{1, l}; \\ m = \arg \min x_i, i = \overline{1, l}; \end{cases} \quad (2.1)$$

где  $x_i$  – матрица вероятностей для цепочки гена новой особи;

$l$  – размер матрицы  $x_i$ ;

После нахождения элементов с минимальной и максимальной вероятностей в массиве, запоминаются их индексы. На основе полученных индексов происходит обмен генов.

Следующим этапом является мутация полученных генов. Мутация приводит к появлению особей с новыми признаками, отличающимися от родительских особей. При этом количество мутаций не должно быть большим, с целью сохранения структуры генотипа популяции.

Сформированная цепочка генов, приходит процедуру мутации, по формуле 2.2. После мутации получается новая цепочка.

$$p(s \rightarrow s^0) = (1 - \mu)^{l-|s|} * \mu^{|s|} \quad (2.2)$$

где  $|s|$  – количество 1 в строке генов;

$\mu$  – вероятность превращения 1 в 0.

В существующих методах, параметр  $\mu$  указывается пользователем или создается программой случайным образом. В разработанном методе параметр  $\mu$  определяется весом нейронной сети для каждого гена, таким образом  $\mu$  изменяется с каждым шагом, а не остается статически заданным

Наиболее эффективна мутация, которая затрагивает два гена из всей цепочки, иначе, существует вероятность того, что потомки будут отличаться от своих родителей, а алгоритм не будет способен обучаться на полученных данных. Алгоритм проходит по всей цепочке, определяет наиболее вероятные гены для мутации и инвертирует его значение ( $0 \rightarrow 1$  или  $1 \rightarrow 0$ ). В случае если значения генов малы, то они меняются местами.

Первый метод основывается на процедуре обмена генов строки между собой. На рисунке 2.4 показан начальный набор генов до перестановки.

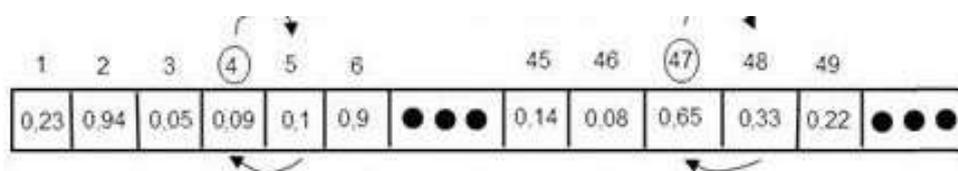


Рисунок 2.4 – Гены до мутации

На рисунке 2.4 показана цепочка хромосом с использованием метода перестановки генов.

Второй метод использует функцию мутации генов, после мутации генов получается новая цепочка, показанная на рисунке 2.5.

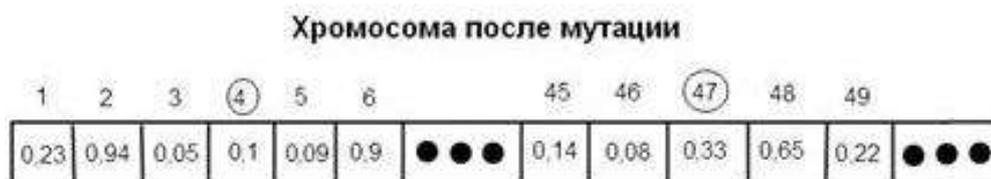


Рисунок 2.5 – Ген после мутации

Использование нескольких операторов мутации позволяет осуществить эффективный поиск списка новых особей, удовлетворяющего требованиям. Это позволяет в короткие сроки получать результаты. Алгоритм будет работать до тех пор, пока не будет достигнуто условия остановки алгоритма. Дополнительным условием остановки является не изменчивость параметра выживаемости с каким-то значением погрешности, при этом алгоритм не будет заходить в бесконечный цикл. По окончании формирования новой популяции идет ее ранжирование согласно по параметру выживаемости, таким образом получается список где на первом месте стоят наиболее приспособленные особи.

Алгоритмическая реализация первого метода без использования нейронных сетей:

- 1) выбираются две особи (родители);
- 2) находится место, где хромосома делится пополам;
- 3) слияние двух частей код, полученных от первого и второго родителя.
- 4) обмен генов местами.

На основании предложенного метода моделирования процесса скрещивания родительских особей с обменом генов без использования нейронной сети был разработан алгоритм, блок-схема которого представлена на рисунке 2.6.

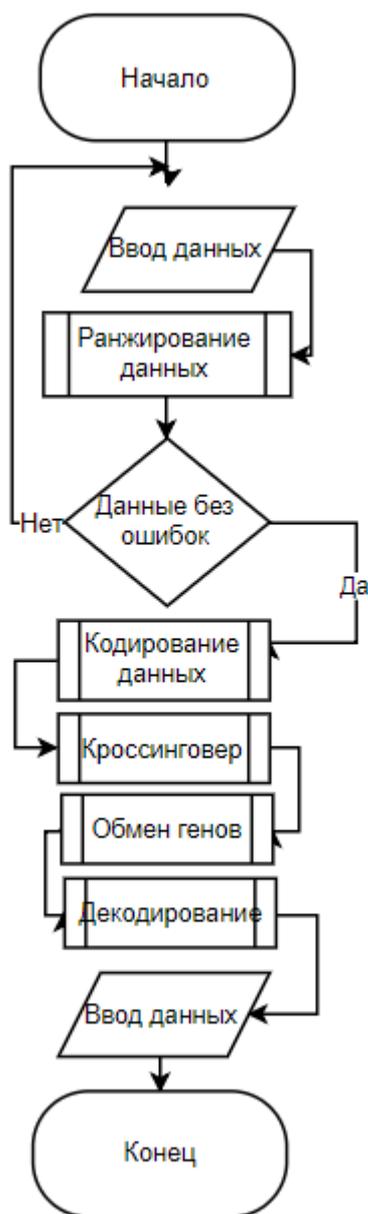


Рисунок 2.6 – Схема алгоритма моделирования процесса скрещивания родительских особей с обменом генов

Первый метод, производит выборку двух родительских особей с наилучшими показателями, переводит цифровые коэффициенты в двоичный код, разбивает цепочки пополам и смешивает их. После чего происходит обмен геном, случайным образом выбираются два гена из всей цепочки и меняются местами. Затем происходит декодирование новых особей, а результат выдается специалисту.

На основе полученных результатов формируется алгоритм для второго метода моделирования процесса селекции с использованием нейронных сетей:

- 1) Задач начальные значения;
- 2) Задание контрольных значений
  - а) Ранжирование;
  - б) Разделение на группы;
  - в) Формирование соотношений;
- 3) Пока генетический алгоритм не сходится выполнять:
  - а) Отобрать родительских особей;
  - б) Скрещивание родительских особей;
  - в) Случайным образом инвертировать участки полученной цепи генов;
  - г) Оценить возможность выживаемости;
  - д) Разделить и записать полученный результат;
  - е) Декодировать полученный результат;
- 4) Оценить полученный результат;
- 5) Выдача результатов на основе нейронной сети;
  - а) Оценка специалистом.

На рисунке 2.7 представлен метод моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети, обучаемый с помощью генетического алгоритма.

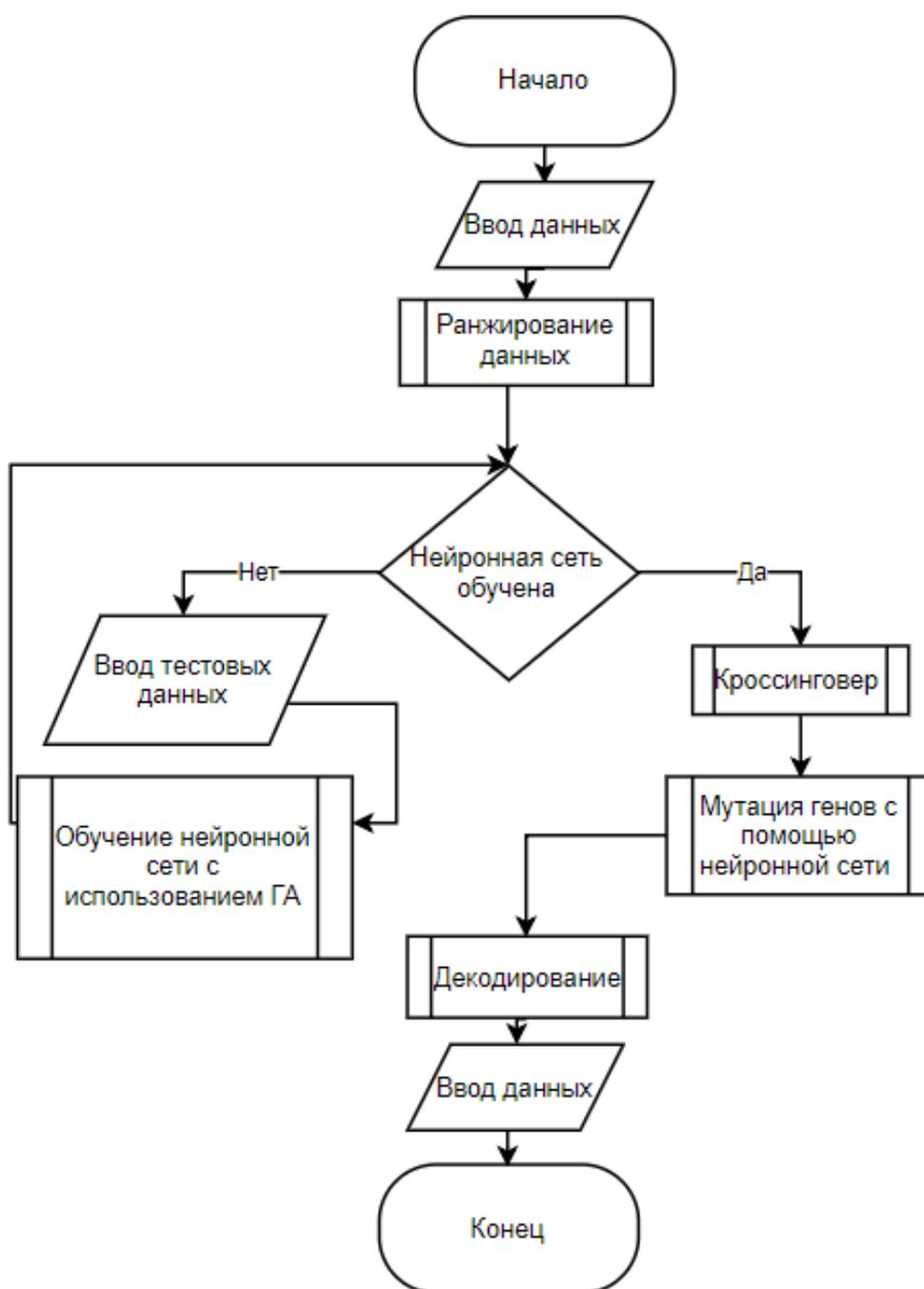


Рисунок 2.7 – Схема алгоритма моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети

Второй метод работает на основе нейронной сети. Сперва необходимо обучить нейронную сеть, сформировав тестовую выборку. За обучение нейронной сети отвечает генетический алгоритм.

Данный метод, производит выборку двух родительских особей с требуемыми показателями, смешивает гены, кодирует их с помощью кода

Грея, далее используется нейронная сеть для участия в процессе мутации генов (формула 2.1). После чего, полученные данные декодируются и выдаются специалисту.

Получив конечную цепочку генов, заносим ее в таблицу с новым поколением, предварительно разделив всю цепочку на несколько маленьких цепочек по 4 символа. Затем декодируем полученные значения согласно таблице 2.2 или переводим из двоичного в десятичный код. Проводим анализ полученных числовых значений по каждому параметру, делаем выводы.

## 2.2 Проектирование модели решения задач селекции

Прежде чем переходить к процессу моделирования, необходимо разобраться с объектом моделирования и на его основе разработать алгоритм. Математическая модель — это приближенное описание объекта моделирования, выраженное с помощью математической символики. Далее необходимо рассмотреть доступное программное обеспечение и методы, на основе которых они работают.

Чтобы новые производственные сорта обеспечивали стабильно высокие урожаи, необходимо в процессе их выведения оценивать селекционный материал признакам.

Все показатели сложны в оценке, необходимо учитывать условия выращивания. Каждый год и на любой стадии селекционного процесса необходимо точно знать, чем вызваны различия в урожайности и качестве продукции разных сортов и селекционных номеров. Рассмотренные методы скрещивания являются разновидностью селекции путем соединения двух генотипов через оплодотворение [36].

При рассмотрении сортов растений необходимо выделить несколько категорий: растения, способные выживать в нормальных условиях; растения

способные выживать в аномальных условиях. Для каждой из категорий, выдвигается список требований, по которым происходит отбор.

Исходя из параметров для селекции растений, доступных из литературных источников, можно сделать вывод, что наиболее важным из них является приспособляемость нового организма к внешним условиям и дача потомства. В рамках реализации поставленной задачи, дополнительными параметрами станут: количество плодов с одной единицы растения, устойчивость к паразитам, вредителям и болезням.

Признаки, по которым оценивают селекционный материал в процессе выведения сортов, делятся на две группы: прямые и косвенные. Оценка сортов по прямым признакам дается путем подсчета, взвешивания, измерения. Например, число сохранившихся к весне растений — прямой показатель зимостойкости. К косвенным признакам относятся биохимические и технологические показатели, среди которых процент и прочность клейковины.

После чего производятся расчеты, которые призваны минимизировать затраты во время проведения эксперимента. Наиболее подходящим математическим методом является генетический алгоритм, который представляет собой способ исследования природных явлений, оптимизации решения задач и моделирования путем подбора исходных параметров и математического аппарата. Математические модели строятся для подбора растений с более высоким шансом выживаемости в реальных условиях.

Генетический алгоритм служит для обучения нейронной сети, то есть подбора таких значений весов, при которых достигается наилучший результат скрещивания. В его основе лежит метод случайного поиска. Недостатком метода является неопределенность во времени для решения задачи. Для того, чтобы генетический алгоритм работал корректно, необходимо представить каждого родителя набором качеств, которые достоверно характеризует его.

В общей схеме генетического алгоритма используется модель оптимизации, представленная формой 2.3 или формулой 2.4:

$$x' = \arg \max W(x) \quad (2.3)$$

$$x' = \arg \min W(x) \quad (2.4)$$

где  $x'$  – элемент нового поколения;

$W(x)$  – матрица начальных значений.

В классическом генетическом алгоритме начальная популяция формируется случайным образом, после чего происходит фиксация размерности ( $n$ ), т.е. количество особей в популяции. Для выполнения поставленных целей, каждый шаг будет выполняться пользователем, который на вход «подает» заранее сформированную начальную популяцию.

Исходя из вышеизложенного, математическая модель будет иметь формульный вид, представленный по формуле 2.5:

$$z = \sum_{i=1}^n c_{ij} x_{ij} \rightarrow \max, \quad j = 1..m \quad (2.5)$$

Ограничения для математической модели генетического алгоритма рассчитываются по формуле 2.6:

$$\begin{aligned} \sum_{j=1}^m c_j &= 1 \\ \sum_{i=1}^n \sum_{j=1}^m x_{ij} &\geq 0 \end{aligned} \quad (2.6)$$

где  $c_{ij}$  – значение параметра начальной популяции;

$x_{ij}$  – количество представителей данного предка.

Алгоритм будет работать до тех пор, пока не выполняются условия ограничений (формула 2.6) при этом целевая функции (формула 2.5) будет оптимальной. Математическая сложность заключается в нахождении среднего значения целевой функции и среднеквадратичного отклонения, так как начальная выборка может содержать большое количество входных данных. Эта задача относится к линейному типу математического программирования.

С целью нахождения оптимальных значений  $x_{ij}$  воспользуемся одним из методов линейного программирования.

### 2.3 Разработка программного интерфейса

Создав два метода и разработав алгоритм их работы, а также модифицирования математической модели работы генетического алгоритма переходим к разработке прототипа программного обеспечения, реализующего представленные методы. Для написания нейронной сети выберем среду программирования MathLab, представленного на рисунке 2.8.



Рисунок 2.8 – Начальная страница MatLab

Представленная среда разработки может реализовать следующие типы нейронных сетей: однослойный перцептрон; многослойный перцептрон; сеть Хэмминга; сеть Ворда; сеть Хопфилда; сеть Кохонена; когнитрон Неокогнитрон. На рисунке 2.9 показана разработанная нейронная сеть с использованием среды разработки MathLab 2014.

```

%% 1) Загрузка исходных данных
% 1.1) Загрузка исходных данных из файла VOLUMES_EUR.mat
load VOLUMES_EUR VOLUMES_EUR;
Data = VOLUMES_EUR;

% 1.2) Указание периодов обучения сети и периода тестирования
TrainStartDate = datenum('01.04.2018', 'dd.mm.yyyy');
TrainEndDate = datenum('30.04.2018', 'dd.mm.yyyy');

TestStartDate = datenum('01.05.2018', 'dd.mm.yyyy');
TestEndDate = datenum('31.15.2018', 'dd.mm.yyyy');

%% 2) Предварительная обработка исходных данных
% 2.1) Масштабирование данных по потреблению: [min - max] -> [0 - 1]
Min = min(Data(:,3));
Max = max(Data(:,3));
B = Min; % Величина сдвига
K = 1/(Max - Min); % Коэффициент масштабирования
Data(:,4) = (Data(:,3) - B) .* K; % Масштабирование

% 2.2) Получение массива для обучения сети P: h, h-24; h-48 (см архитектуру сети)
Index0 = find(Data(:,1) >= TrainStartDate & Data(:,1) <= TrainEndDate);
Index1 = find(Data(:,1) >= TrainStartDate-1 & Data(:,1) <= TrainEndDate-1);
Index2 = find(Data(:,1) >= TrainStartDate-2 & Data(:,1) <= TrainEndDate-2);
P(:,1) = Data(Index0,4);
P(:,2) = Data(Index1,4);
P(:,3) = Data(Index2,4);

%% 3) Настройка нейронной сети
% 3.1) Описание архитектуры нейронной сети
NumberOfInputNeurons = 48; % Количество входов нейронной сети
NumberOfHiddenNeurons = 72; % Количество нейронов в скрытом слое
NumberOfOutputNeurons = 24; % Количество выходов

% 3.2) Описание параметров нейронной сети
LearningRate = .1;
Momentum = .5;
Epochs = 300;
Wli = 0;
bli = 0;

```

Рисунок 2.9 – Код нейронной сети

Показанный код был написан конструктором в визуальной среде разработки.

На рисунке 2.10 показан интерфейс программы, написанный на языке объектно-ориентированного программирования C#.

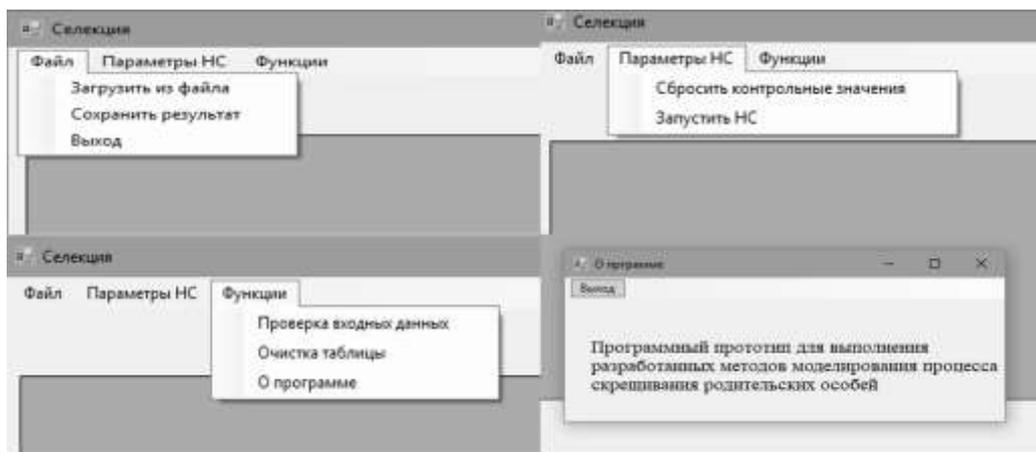


Рисунок 2.10 – Меню программы

Созданная нейронная сеть была экспортирована в C# используя встроенные библиотеки MATLAB ComponentRuntime. Программа не требует установки, но необходимо соблюдение расположения внутренних элементов, поскольку в программе реализована возможность чтения исходных данных из файла.

Запуск нейронной сети осуществляется через верхнее меню. После этого активируется процесс обучения нейронной сети на основе генетического алгоритма. Генетический алгоритм запускается самостоятельно в коде программы. Для запуска метода с обменом генов достаточно нажать кнопку «Рассчитать»

## 2.4 Основные результаты и выводы

На основании проведенного анализа методов селекции, были разработаны методы моделирования процессов скрещивания родительских особей. Из теоретического анализа генетического алгоритма, была изменена его математическая модель под нужды разработанных методов. Проведена разработка прототипа программного обеспечения для решения задач селекции. Стоит отметить, что несколько способов мутации позволяют сократить время обучения и обеспечить качество генов новой особи.

### 3 Проведение вычислительных экспериментов

#### 3.1 Оценка эффективности разработанных методов

В данном разделе будет проведена апробация на практике разработанного метода селекции, в учреждении, которое занимается ведением статистики и учетом сельскохозяйственной деятельности всех организаций, находящихся на территории муниципального района. Будет приведена статистика сбора урожая по нескольким районам, а также сформирована общая выборка (обучающая и контрольная).

Разработанный метод и программное обеспечение, выполняющее его, было установлено на центральный сервер муниципальной администрации Волоконовского района. Сервер установлен в защищенном месте и имеет разделение прав доступа к хранящейся на нем информации.

Выбор места внедрения обусловлен высокими показателями итогами развития за прошлый год в области сельского хозяйства и животноводства. Ведущей отраслью в районе является сельское хозяйство. В производстве сельскохозяйственной продукции на 1 га пашни Волоконовский район занимает второе место.

Этапы проведения апробации разработанного метода:

- 1) прибытия в организацию;
- 2) запрос статистических данных;
- 3) подготовка общей выборки;
- 4) установка программного обеспечения для решения задач селекции;
- 5) настройка удаленного доступа к серверу и настройка пользователей;
- 6) запуск и настройка нейронной сети на обучающей выборке;
- 7) контрольное тестирование;

8) анализ результатов.

Апробация на тестовой выборке будет проходить следующим образом: в программу «подаются» данные за 2016 год, результаты, полученные человеком и машиной, сравниваются.

После запуска программы необходимо задать начальные значения, это можно делать вручную, можно подгрузить из файла. На рисунке 3.1 показаны уже загруженные начальные данные.

F1	1	2	3	4	5
Длина стебля	12	10	8	11	2
Плобовитость	10	10	9	15	13
Выживаемость	1	2	4	8	6
Необходимость ...	7	8	3	5	11
устойчивость к ...	0	1	5	5	7

Рисунок 3.1 – Начальные данные

На рисунке 3.2 показаны результаты тестирования первого метода, скрещивания родительских особей с обменом генов, переводит цифровые коэффициенты в двоичный код, разбивает цепочки пополам и смешивает их

F1	1	2	3	4
Длина стебля	3	2	2	3
Плобовитость	0	3	3	0
Выживаемость	2	3	3	2
Необходимость ...	1	2	1	3
устойчивость к ...	0	2	0	1

Рисунок 3.2 – Результат работы первого метода

На рисунке 3.3 показаны результаты применения второго метода моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети, кодирует их с помощью кода Грея, после чего используется нейронная сеть для процесса мутации генов

F1	1	2	3	4
Длина стебля	13	9	12	8
Плобовитость	7	10	4	11
Выживаемость	4	1	1	5
Необходимость ...	10	6	8	4
устойчивость к ...	2	2	1	0

Рисунок 3.3 – Результат работы второго метода

Полученные цифры обозначают качественные характеристики новых особей по выбранным параметрам. Результаты работы второго метода оказались лучше, результатов работы первого метода. Подобный процесс занимает в других программных решениях несколько дней, так как им необходимо накопление статистических данных изменений за прошлые поколения растений.

На рисунке 3.4 показаны закодированные данные по алгоритму Грея, исходя из особенностей среды, первые нули, система не отображает.

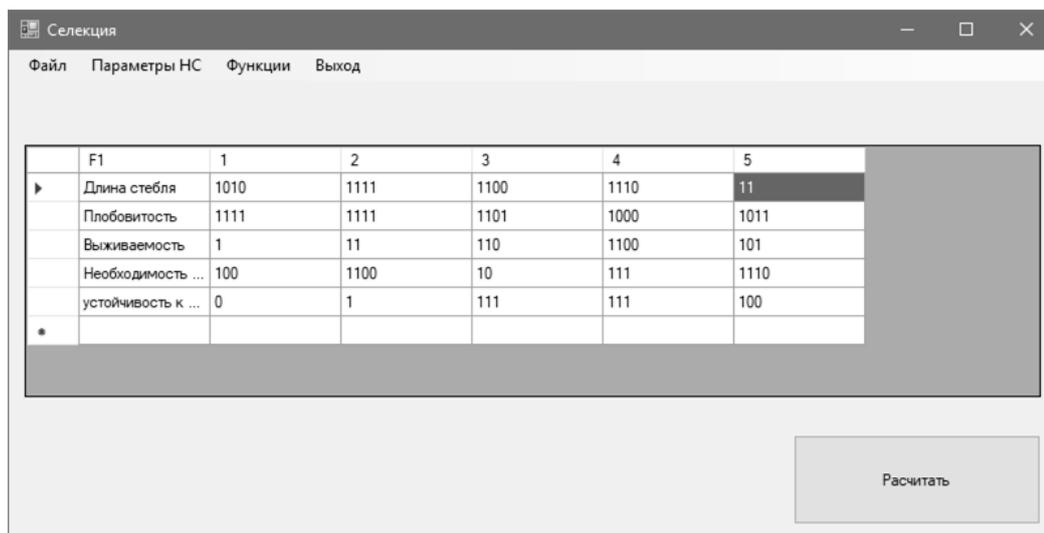


Рисунок 3.4 – Кодировка по Грею

Данная особенность среды программирования учтена при получении конечного гена родительской особи. Закодировав значения параметров, все значения проходят процедуру конкатенации по вертикали и ранжируются по ключевому признаку. Особь, которая осталась без пары, скрещивается в первой обобью в таблице. Созданные пары представляют собой значения всех признаков в виде строки. К данной строке применяются процедуры разработанных методов. Процедура мутации, где каждому гену закодированной хромосомы присваивается случайное значение из интервала  $[0, 1]$ . Процедура кроссинговера, где две строки делятся случайным образом пополам и скрещиваются. Процедура обмена – обмен генов местами.

Процедура кодирования характеристик параметров растений с использованием двоичного преобразования, представлена на рисунке 3.5.

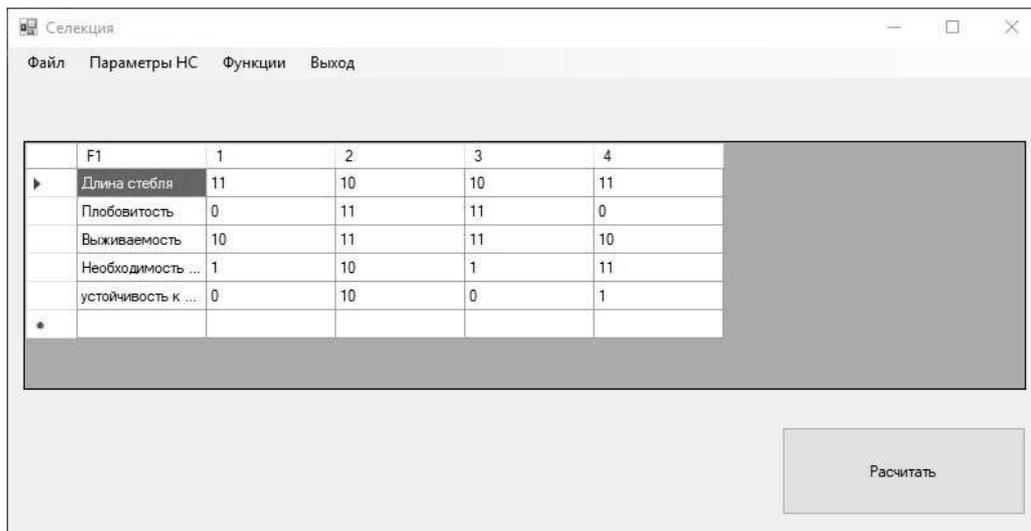


Рисунок 3.5 – Двоичное кодирование

Исходя из результатов, можно сделать вывод, почему второй метод работает гораздо лучше, так как использует кодирование данных по методу Грея. Поскольку метод Грея кодирует каждое число маской из четырех цифр, а двоичное кодирование нет, то теряется часть параметров при скрещивании генов и при дальнейшем декодировании.

В программе предусмотрена возможность сброса весовых коэффициентов нейронной сети. Если результаты значительно отличаются от ранее полученного, т.е. нейронная сеть переобучалась, то веса нейронной сети сбрасываются, на рисунке 3.6 показана функция, для выполнения сброса.

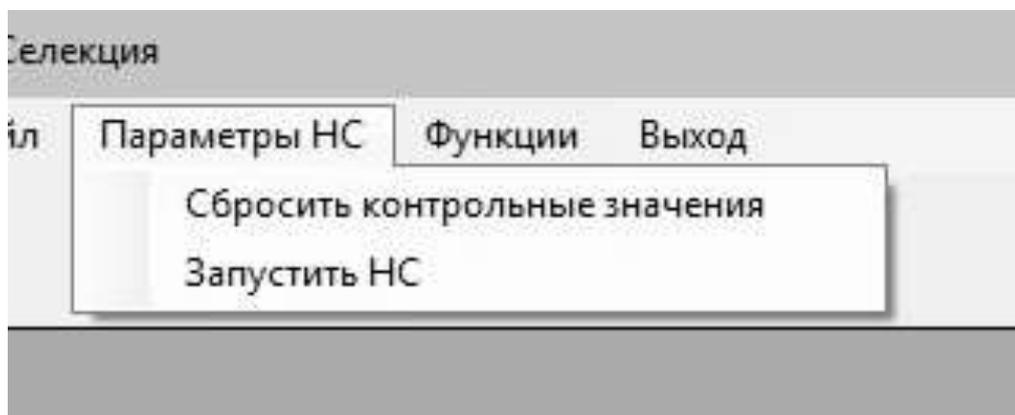


Рисунок 3.6 – Сброс весов нейронной сети

На рисунке 3.6 показана функция («Сбросить контрольные значения»). Активация этой функции подразумевает замена весовых коэффициентов на случайные числа.

Закончив процесс обучения по данным за 2016 год, осуществляется переход на контрольные данные. Полученные данные на этом этапе апробации можно применять на практике в реальном учреждении.

В Волоконовском районе выращивается большое количество культур, и занимают значительную часть посевных площадей, примерно 1130,6 тыс. гектар.

Тестовые данные были предназначены для первичной настройки весов, теперь необходим реальный набор данных (рисунок 3.7), который отвечает требованиям организации.

	Урожайность	Устойчивость к болезням	Необходимость полива	Длина стебля	Выживаемость
Озимые зерновые	324	34	2	65	390
Пшеница	325	34	2	72	397
Тритикале	17	1	0	1	17
Кукуруза	125	14	2	17	144
Овес	6	0	0	4	10
Гречиха	6	2	0	4	10
Горох	5	1	0	2	8
Свекла	67	2	1	6	74
Картофель	1	1	41	0	42

Рисунок 3.7 – Данные апробации

На рисунке 3.7 представлена выдержка из всего перечня контрольных данных и содержит как индивидуальные показатели, так и общие показатели по группам растений.

Представленные данные будут использоваться для сравнения второго метода с уже существующими методами. В данных содержатся и общие показатели урожайности за год, данные с другими видами растений. Для более полной настройки весов, был сформированный тестовый набор данных,

представлен в таблице 3.1 и представляет собой зерновые культуры, выращиваемые в районе.

Таблица 3.1 – Набор данных апробации

Показатель \ Родитель	Пшеница 1	Пшеница 2
Зрелость зерен	34	12
Испорченность	1	1
Окрас	64	8
Морозобойность	123	24
Влажность	125	140
Засоренность	97	101

В выборке были значения не только о зерновых, но и значения зернобобовых культур, масленичных культур, картофель и другие представители сельского хозяйства. Всего, во время тестирования и настройки весов нейронной сети, было использовано 215 входных параметров (рисунок 3.7). Общая выборка была поделена следующим образом: 70% наблюдений выборки использовалось для обучения генетического алгоритма, 30% использовалось в качестве контрольной выборки для оценки настройки весовых коэффициентов обученной нейронной сети генетического алгоритма.

Все данные вносились в таблицу Excel и загружались в программу. Результат применения второго метода моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети по входным данным из таблицы 3.1, представлен на рисунке 3.8.

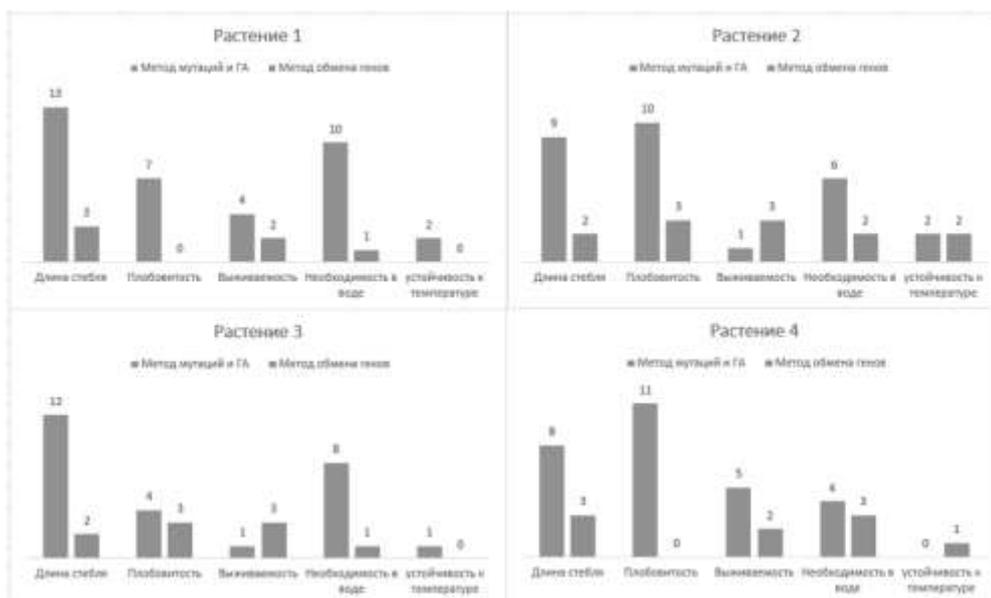


Рисунок 3.8 – Итоги скрещивания

На рисунке 3.8 демонстрируются итоги скрещивания двух родительских особей из таблицы 3.1. В алгоритме стоит ограничение, на каждую пару особей не более 4 новых особей.

Так как аналогичных программных методов решения задач селекции на момент написания работы было найдено немного, то сравнительная характеристика полученного результата проводилась с результатами уже существующих решений, использующих метод VLUP. На рисунке 3.9 показаны результаты работы программы на основе метода VLUP и разработанного программного решения на основе второго метода.

Результаты метода VLUP									Результат разработанного метода				
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000	F1	Особь1	Особь2	Особь3	Особь4
0,000	1,167	0,167	-0,500	-0,667	-0,167	0,000	0,167	1,000	Зрелость зерна	53	40	57	30
0,000	0,167	1,315	0,093	-0,074	-0,315	-0,037	0,019	-1,000	Испорченность	53	43	43	29
0,000	-0,500	0,093	0,870	0,370	-0,093	-0,148	-0,093	-1,000	Скорос	26	23	27	20
0,000	-0,667	-0,074	0,370	1,037	0,074	-0,148	-0,259	-1,000	Морозобойность	115	96	105	27
0,000	-0,167	-0,315	-0,093	0,074	0,315	0,037	-0,019	0,000	Влажность	46	40	53	39
0,000	0,000	-0,037	-0,148	-0,148	0,037	0,259	0,037	0,000	Засоренность	46	41	53	33
0,000	0,167	0,019	-0,093	-0,259	-0,019	0,037	0,315	0,000					
1,000	1,000	-1,000	-1,000	-1,000	0,000	0,000	0,000	0,000					

Рисунок 3.9 – Сравнительная таблица

Результаты, показывают, что числовые коэффициенты характеристик для нового поколения особей близкие к нулю и значительно отличаются от начальных данных в худшую сторону. Числовые показатели разработанного решения показывают хорошие результаты, приближенные к реальным, что говорит о работоспособность решения. Плохие числовые показатели метода BLUP говорят о необходимости добавления данных тех же особей за предыдущие года.

Подытожив работу, проведенную по апробации метода на практике в целях упрощения решения задач селекции, был разработан ряд предложений и рекомендаций:

- 1) улучшение способов формирования данных характеризующих растения;
- 2) совершенствование структуры алгоритма;
- 3) учесть в программе возможность обучения нейронной сети с помощью учителя;
- 4) создать более узкую структуру алгоритма;

### 3.2 Основные результаты и выводы

На основании проведенных вычислительных экспериментов удалось определить наилучший метод из представленных, так же проведено сравнение данного метода с существующим методом BLUP. На основе полученных результатов экспериментов были разработаны рекомендации по улучшению и применению разработанных методов. Внедрение прототипа программного обеспечения позволит улучшить показатели и повысить количество собираемого урожая с 1 га пашни. Для работы программного обеспечения были настроены права доступа к программе, подготовлены тренировочные и тестовые данные, на основании которых проходило обучение и тестирование.

## ЗАКЛЮЧЕНИЕ

Моделирование процессов играет важную роль в создании или испытании новых методов или алгоритмом, позволяя достоверно определить поведение реального объекта в идентичных условиях. Использование моделирования и достижений в области вычислительных технологий способствует ускорению процесса селекции. Поэтому очень важно изучить информационные технологии при моделировании процессов скрещивания растений. В этих целях было принято решение по созданию методов и алгоритма моделирования процесса селекции с использованием нейронных сетей, которые позволят ускорить процесс решения задач селекции.

В первом разделе выпускной квалификационной работы были произведены исследование и анализ предметной области, анализ методов селекции и состояния информационного обеспечения процессов моделирования. Анализ уже существующих разработок показал, что существует малое количество программ, позволяющих решать задачи селекции. Развитием селекции занимаются с теоретической стороны. Селекция сложный эволюционный процесс, который был разработан человеком для выращивания культур с необходимым набором качеств. В результате исследования предметной области была поставлена задача по созданию методов моделирования процесса селекции.

На основании оценки существующих методов решения задач селекции, были рассмотрены эксперименты и результаты внедрения нейронных сетей в различные сферы. Рассмотрены методы обучения нейронных сетей, структур. Для использования нейронных сетей необходимо форматирование входные данных и обучить ее.

Для наиболее правильной разработки метода с использованием нейронных сетей, были проанализированы существующие структуры нейронных сетей, проанализированы примеры их успешного применения и

способы обучения в различных сферах деятельности. Выделены требования для родительских особей, сформирован список качеств и модель представления данных в программе, а также вывод информации пользователя. Были разработаны процедуры кодирования, мутации, обмен генов и кроссинговер родительских особей. На основании проведенного исследования проблемы моделирования и процессов селекции были разработаны два метода решения задач селекции: метод скрещивания родительских особей с обменом генов без использования нейронной сети и метод моделирования скрещивания родительских особей с учетом мутаций с использованием нейронной сети, которые позволят ускорять процесс моделирования селекции. Разработка системы производилась в среде программирования MathLab 2014 и C#.

Разработанная система прошла тестирования, в результате которого доказала свою работоспособность. При тестировании системы использовалась тестовая выборка, предоставленная управлением сельского хозяйства.

Основные результаты и выводы проделанной работы:

1) проанализированы методы и программное обеспечение методов, позволяющие решать задачи селекции с использованием информационных технологий;

2) разработаны методы и алгоритмы для решения задач селекции, позволяющие моделировать процесс скрещивания родительских особей, разработаны процедуры и функции, которые отличаются от существующих, применением нейронной сети для решения задач селекции;

3) разработан прототип программной реализации разработанных алгоритмов, позволяющий использовать оригинальную модель представления данных;

4) проведены вычислительные эксперименты, показавшие работоспособность разработанных методов;

5) проведено сравнение результатов применения разработанных методов и метода BLUP, иллюстрирующее преимущество разработанных методов.

Для решения поставленных задач было изучено несколько существующих практических разработок для оценки селекционного потенциала животных и оценки питательной смеси. В процессе написания выпускной работы были выполнены поставленные задачи:

- 1) проанализированы методы и программное обеспечение моделирования решения задач селекции;
- 2) проведен анализ нейронных сетей и возможностей их применения;
- 3) разработан метод моделирования процесса скрещивания родительских особей с обменом генов без использования нейронной сети;
- 4) разработан метод моделирования процесса скрещивания родительских особей с учетом мутаций с использованием нейронной сети;
- 5) разработано программно – алгоритмическое обеспечение;
- б) проведены вычислительные эксперименты.

Во время проведения экспериментов удалось достичь поставленной цели. Апробация программного обеспечения показала, что разработанные методы работоспособны на практике. Все показатели можно улучшить, если за обучение нейронной сети будет отвечать специалист в этой области.

Программное решение разрабатывалось для моделирования селекции в растениеводстве и сельском хозяйстве, но также может быть использована для моделирования процессов селекции в животноводстве.

Рассмотренные особенности позволили выделить основные аспекты деятельности селекционера и сформулировать основные требования, выдвигаемые растениям родительских особей, новому поколению особей, процессу моделирования, мутации и скрещиванию. Применения системы на практике позволит оптимизировать и ускорить процесс решения задач селекции.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Andrzej, C. Neural Networks for Optimization and Signal Processing [Текст]/ Andrzej Cichocki, R. Unbehauen, John Wiley and Sons Ltd, 1993. – 526 с
- 2 Fredric, M. Ham. Principles of neurocomputing for science and engineering [Текст]/ McGraw-Hill Science, 2000. – 642 с.
- 3 Harold, Abelson. Structure and Interpretation of Computer Programs [Текст]/ Harold Abelson, Jerry Sussman, Julie Sussman MIT Press, 1984 – 588 с.
- 4 L. Chambers. The practical handbook of genetic algorithms. Applications. vol. [Текст]/ Chapman and Hall/CRC Reference, 2001. – 544 с.
- 5 Rumelhart, B.E. Learning representations by back propagating error [Текст]/ Rumelhart B.E, Minton G.E., Williams R.J. Wature, 1986. – 323 с.
- 6 Teuvo, Kohonen, Self-Organized Formation of Topologically Correct Feature Maps [Текст] /Department of Technical Physics, Helsinki University of Technology, Espoo, Finland,1982. – 11 с.
- 7 Кесман, В. Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models [Текст] /MIT Press, 2001. – 541 с.
- 8 Аверкин, А. Нечеткие множества в моделях управления и искусственного интеллекта [Текст]/ Аверкин А., Батыршин И., Блишун, В. Силов А., Тарасов В. – М.: Наука, 1986. – 183 с.
- 9 Александр, Галушкин, Нейрокомпьютеры. Учебное пособие [Текст]/ Альянс, 2014. – 528 с.
- 10 Белов, К. Д.. Многослойная нейронная сеть [Текст]/ ВНИИМТ, 2011. – 60 с.
- 11 Бокс, Дж. Анализ временных рядов. Прогноз и управление [Текст]/ Бокс Дж, Дженкинс Г. М.: Мир, Вып.1, 1974. – 406 с.
- 12 Бокс, Дж. Анализ временных рядов. Прогноз и управление [Текст]/ Бокс Дж, Дженкинс Г. М.: Мир, Вып.2., 1974. – 197 с.

- 13 Бороевич, С. Принципы и методы селекции растений. [Текст]/ – Москва: Колос, 1984. – 344 с.
- 14 Валентина, Грязева. Селекция растений. [Электронная книга]/ – Пенза, 2012. – 190 с.
- 15 Васильев, А. Принципы и техника нейросетевого моделирования [Текст]/ Александр В., Дмитрий Т, // Нестор-История, 2014. – 218 с.
- 16 Виноградов, А. Н. Выделение и распознавание локальных объектов на аэрокосмических снимках. [Текст]/ Виноградов А.Н., Калугин Ф.В., Недев М.Д., Погодин С.В., Талалаев А.А., Тищенко И.П., Фраленко В.П., Хачумов В.М. – Авиакосмическое приборостроение № 9, 2007. – 39-45 с.
- 17 Владимир, Р. Эволюция. Нейронные сети. Интеллект. Модели и концепции эволюционной кибернетики [Текст]/ Едиториал УРСС, 2017. – 224 с.
- 18 Вороновский, Г. К. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. – Х.: основа, 1997. – 483 с.
- 19 Гаврилова, Т. А. Базы знаний интеллектуальных систем. Учебник. [Текст]/ Гаврилова Т. А., Хорошевский В. Ф. — СПб.: Питер, 2000. – 384 с.
- 20 Галушкин, А. И. Нейроматематика (проблемы развития). [Текст]/ – М.: Радиотехника, 2003. – 40 с.
- 21 Гиляров, М. С. «Биологический энциклопедический словарь.» Гл. ред. [Текст]/ М. С. Гиляров; Редкол.: А. А. Бабаев, Г. Г. Винберг, Г. А. Заварзин и др. — 2-е изд., исправл. — М.: Сов. Энциклопедия, 1986. – 831 с;
- 22 Горелик, А. Л. Селекция и распознавание на основе локационной информации. [Текст]/ Горелик А.Л., Барабаш Ю.Л. – Букинист, 1990. – 240 с.
- 23 Гуляева, Ю. В. Нейрокомпьютеры в системах обработки сигналов [Текст]/ Гуляева Ю. В., Галушкина А. И., – М: Издательство «Радиотехника», 2003. – 176 с.
- 24 Джеффри, Е. Хинтон. Как обучаются нейронные сети. [Текст]/ – В мире науки ,1992. – с. 103-107.
- 25 Джеффри, Е. Хинтон. Как обучаются нейронные сети. [Текст]/ В мире науки N 1, 1992. – 103-107 с.

26 Джеффри, Е. Хинтон. Нейронные сети. STATISTICA Neural Networks. [Текст]/ М.: Горячая линия– Телеком, 2000 г. – 182 с.

27 Донской, Д. А. Применение аналитических технологий в системах управления и информатике [Текст]/ Донской Д.А., Слепцов Н.В., Щербаков М.А. – Пенза, 2005.

28 Ефименко, Г. А., Сеница А.М.: Нейронные сети в MatLab [Электронный ресурс] // Digiratory. 2017 г. URL: <https://digiratory.ru/508> Зюзысов В. М. Математическое введение в декларативное программирование: учебное пособие. — Томск: ТГУ, 2003. — 83 с.

29 Крисиллов, В. А. Применение нейронных сетей в задачах интеллектуального анализа информации. [Текст]/ Крисиллов В. А., Олешко Д.Н., Трутнев А.В. – Труды Одесского политехнического университета, Вып.2 (8), 1999. – 134 с.

30 Маилян, А. Л. Распознавание объектов в сложных иерархических системах при оптимизации рисков строительных объектов. [Текст]/ Маилян А. Л., Янин А. Г. – Воронеж: Воронежский государственный архитектурно-строительный университет, 2014. – 178-182 с.

31 Мак-Каллок, У. С. Логическое исчисление идей, относящихся к нервной активности [Текст]/ Мак-Каллок У. С., Питтс В., Шеннона и Дж. Маккарти. — М.: Изд-во иностр. лит., 1956. — 363-384 с.

32 Мищенко, В. А. Использование генетических алгоритмов в обучении нейронных сетей [Текст]/ Мищенко В.А., Коробкин А.А.: Современные проблемы науки и образования № 6. 2011. – 280 – 288 с.

33 Нечаева, Ю. И. Нейрокомпьютеры в интеллектуальных технологиях XXI века [Текст]/ – М: Издательство «Радиотехника», 2012. – 352 с.

34 Племяшов, К. В. Использование метода BLUP Animal Model в определении племенной ценности голштинизированного скота Ленинградской области [Текст]/ К.В. Племяшов, В.В. Лабинов, Н.Р. Рахматулина, Е.И. Сакса, М.Г. Смарагдов, А.А. Кудинов, А.В. Петрова – Молочное и мясное скотоводство №1, 2016. – 2-5 с.

- 35 Постников, А. Н. Математическое моделирование процесса исполнения бюджета [Электронный ресурс] // Постников, А. Н., Гахова Н. Н. – Агентство международных исследований, январь 2017. URL: [http://ami.im/mnpk-ek\\_4/](http://ami.im/mnpk-ek_4/);
- 36 Постников, А. Программная среда разработки нейронной сети [Электронный ресурс] // Постников, А. Н., Ряснова В. А. – Агентство международных исследований, Май 2017. URL: <https://ami.im/mnpk-148/>;
- 37 Постников, А. Н. О методиках обучения искусственных нейронных сетей в задаче селекции животных [Электронный ресурс] // Совет молодых ученых и студенческое научное сообщество Белгородского государственного национально исследовательского университета, Апрель 2017 года.
- 38 Постников, А. Н. Распознавание изображений графических моделей предметной области в научных текстах [Электронный ресурс] // Агентство международных исследований, апрель 2017. URL: <https://ami.im/mnpk-ek-7/>;
- 39 Роберт, Каллан. The Essence of Neural Networks // Вильямс – 2017, с. 288
- 40 Роберт, Каллан. Основные концепции нейронных сетей // Вильямс – 2003, с. 288
- 41 Ряснова, В. А. Постников А. Н.; О МЕТОДИКАХ РАЗРАБОТКИ ИНФОРМАЦИОННЫХ СИСТЕМ // Агентство международных исследований, Май 2017. URL: <https://ami.im/mnpk-148/>;
- 42 Садовой, А. В., Сотник С. Л. Алгоритмы обучения нейронных сетей будущего. // <http://www.alicetele.com/~sergei/articles/algo/algo.htm>
- 43 Саймон, Хайкин. Neural Networks: A Comprehensive Foundation [Текст]/ – М.: Вильямс, 2016. –1104 с.
- 44 Самигуллина, Н. С. Практикум по селекции и сортоведению плодовых и ягодных культур: Учебное издание. [Текст]/ – Мичуринск: Мичуринский государственный аграрный университет, 2006. – 197 с.
- 45 Сеитова, С. В. Классификация нейросетевых структур. [Текст]/ Сеитова С. В., Чижков А. В. – Известия ЮФУ, Технические науки, 209. – 7 с.

- 46 Сергей, Инге-Вечтомов. Генетика с основами селекции [Текст]/ – Н-Л. 2015. – 720.
- 47 Соловьев, А. М. Однослойная искусственная нейронная сеть с гистерезисной функцией активации. [Текст]/ Соловьев А. М., Семенов М. Е., Кабулова Е. Г. – Воронеж: Воронежский государственный университет, 2013 – 188-191 с.
- 48 Трикоз, Д. В. Нейронные сети: как это делается? [Текст]/ Компьютеры + программы N 4(5). 1993. – 14-20 с.
- 49 Тюрин, Ю. Н.. Статистический анализ данных на компьютере. [Текст]/ Тюрин Ю.Н., Макаров А.А., В.Э.Фигурнова. – М.: ИНФРА-М, 1988. – 528 с.
- 50 Уоссермен, Ф. Нейрокомпьютерная техника. [Текст]/ – М.: Мир, 1992. – 361 с.
- 51 Федеральный закон от 27 июля 2006 года № 149-ФЗ «Об информации, информационных технологиях и о защите информации»
- 52 Флегонтов, А. В. Многопоточное обучение нейронных сетей методом обратного распространения ошибки. [Текст]/ Флегонтов А. В., Фомин В. В., Смирнов Е. М. – Российский государственный педагогический университет им. А.И. Герцена, 2014. – 225-226 с.
- 53 Царегородцев, В.Г. Определение оптимального размера нейросети обратного распространения через сопоставление средних весов синапсов [Текст]/ Материалы XIV Международной конференции по нейрокибернетике Т.2, Ростов-на-Дону, 2005. – 60-64 с.
- 54 Царегородцев, В. Г. Оптимизация предобработки признаков выборки данных: критерии оптимальности [Текст]/ Нейрокомпьютеры: разработка, применение №4. 2005. – 64-67 с.
- 55 Цыпкин, Я. З. Информационная теория идентификации. [Текст]/ – М.: Наука. Физматлит, 1995. — 336 с.