

УДК 004.93'1

НЕЙРОСЕТЕВАЯ ИДЕНТИФИКАЦИЯ НЕИСПОЛЬЗУЕМЫХ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ЗЕМЕЛЬ НА СНИМКАХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ В СИСТЕМЕ QGIS



А.Н. Леванков

Инженер-программист ООО «Code Inspiration Pro», магистрант БГУИР



А.А. Дудкин

Заведующий лабораторией идентификации систем ОИПИ НАН Беларуси, доктор технических наук, профессор

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь.
Объединенный институт проблем информатики НАН Беларуси, Республика Беларусь.
ООО «Code Inspiration Pro», Республика Беларусь.
E-mail: alexlevankou@gmail.com, doudkin@newman.bas-net.by.*

А. Н. Леванков

Окончил Белорусский государственный университет, факультет прикладной математики и информатики. Магистрант БГУИР КСис. Работает в ООО «Code Inspiration Pro» в должности инженера-программиста. Проводит научные исследования в области цифровой обработки изображений.

А. А. Дудкин

Заведующий лабораторией идентификации систем Объединенного института проблем информатики НАН Беларуси, д. т. н., проф. Проводит научные исследования в области цифровой обработки изображений, распознавания образов, систем технического зрения.

Аннотация. В данной работе решается задача идентификации неиспользуемых сельскохозяйственных и залежных (заброшенных) земель на снимках дистанционного зондирования Земли. Исходными данными являются многоспектральные снимки поверхности Земли, полученные при помощи космических спутников Sentinel-2, и набор вегетационных индексов, на основе которых оценивается состояние растительности. Создается обучающая выборка, на основе которой строится нейросетевой алгоритм идентификации, результатом работы которого является карта неиспользуемых земель. Обработка данных выполняется на базе открытой геоинформационной системы QGIS.

Ключевые слова: дистанционное зондирование Земли, многоспектральное изображение, неиспользуемая сельскохозяйственная земля, залежная земля, вегетационный индекс, NDVI, идентификация, искусственная нейронная сеть, геоинформационная система, QGIS, Orfeo Toolbox.

Введение.

Современные технологии автоматизированной обработки больших объемов данных нашли широкое применение в различных областях экономики. Сельское хозяйство, считающееся более традиционной и менее инновационной сферой деятельности, также находит новые решения по оптимизации хозяйствования с помощью информационных технологий. На помощь приходят технологии дистанционного зондирования Земли (ДЗЗ) и автоматизированной обработки полученной информации. Анализ состояния растительного покрова поверхности Земли и идентификация выведенных из севооборота участков пашни (залежных земель) является хорошим примером применения данных технологий. Традиционные способы обследования сельскохозяйственных угодий требуют больших человеческих ресурсов и временных затрат.

Одним из подходов к идентификации залежных земель является анализ снимков участков

земли, полученных до и после приостановки на них хозяйственной деятельности [1]. Исследование показало, что максимальная точность идентификации составляет 54-90% для заброшенных пахотных земель и 50-72% для заброшенных пастбищ в зависимости от количества снимков местности. Данные классифицировались с помощью метода опорных векторов. Недостатком данного подхода является необходимость наличия таких снимков до и после приостановления хозяйства на данном участке поля, между которыми может пройти много лет.

Другой подход к идентификации основан на использовании спектральных характеристик изображений для детектирования и оценки состояния земель [2]. Он позволяет работать с актуальными снимками земной поверхности. Ограничением данного способа является необходимость извлечения из изображений информативных признаков для идентификации. Для этой цели необходимо строить репрезентативные выборки, которые позволяют оценить спектрально-яркостные характеристики залежных земель. Отделить залежные земли от пашни на снимках ДЗЗ можно с использованием метода дискриминантного анализа [3], сущность которого заключается в выделении информационных признаков, позволяющих отнести анализируемый участок к тому или иному классу. Такими признаками могут выступать сезонные величины спектрально-отражательных характеристик, например, вегетационные индексы, при этом точность идентификации достигает 70%. Экспертное дешифрирование снимков высокого разрешения (1–2 м/пиксель) позволяет повысить информативность индексов, однако этот способ требует много времени, особенно когда необходимо обследовать многие тысячи гектар полей. Автоматизированное выявление залежей на основе одиночных снимков, как правило, затруднительно, в связи с тем, что в разные сроки вегетационного сезона отражательные свойства залежей могут быть схожи с аналогичными признаками других угодий. К настоящему времени имеется опыт использования различных типов спутниковых данных для оценки состояния залежей, однако проблема их эффективного картографирования на основе материалов дистанционного зондирования остается во многом открытой.

Вегетационные индексы в комбинации с другими характеристиками, например, цветовыми и текстурными, позволяют повысить точность выделения областей интереса и последующее построение карт сельскохозяйственных полей на основе нейросетевого подхода к распознаванию [4, 5]. В данной работе на основе данных дистанционного зондирования строится обучающий набор данных и нейросетевой алгоритм идентификации – извлечение информативных признаков и идентификация участков земной поверхности выполняется искусственной нейронной сетью, результатом работы которого является карта неиспользуемых земель. Обработка данных выполняется на базе открытой геоинформационной системы (ГИС) QGIS и открытого проекта Orfeo Toolbox.

Постановка задачи идентификации.

Идентификация участков на изображениях земной поверхности базируется на решении двух основных задач: сегментации и выделении контуров участков.

Пусть $f(x, y)$ – дискретная функция, описывающая анализируемое изображение; X – конечное множество точек плоскости, на котором определена функция $f(x, y)$; $S = \{S_1, S_2, \dots, S_k\}$ – разбиение X на k непустых связных областей S_i , где $i = 1, 2, \dots, k$; L_p – предикат, заданный на множестве S и принимающий истинные значения тогда и только тогда, когда любая пара точек (x, y) из каждого подмножества S_i удовлетворяет некоторому критерию однородности этого подмножества.

Сегментацией изображения $f(x, y)$ по предикату L_p называется разбиение $S = \{S_1, S_2, \dots, S_k\}$, удовлетворяющее условиям:

- а) $\bigcup_{i=1}^k S_i = X$;
- б) $S_i \cap S_j = \emptyset$ для любых $i \neq j$;
- в) $L_p(S_i)$ принимает истинные значения для любого i ;
- г) $L_p(S_i \cup S_j)$ принимает ложные значения для любых $i \neq j$.

Отметим, что k областей, полученных в результате сегментации часто различных классов, группируются в m классов, где $2 \leq m \leq k$.

Большинство известных методов сегментации и выделения контуров, можно разделить на две основные группы: областно-ориентированные, когда непосредственно строятся S_i , и гранично-ориентированные, когда определяются границы между S_i [6, 7]. В методах первой группы критерием однородности при задании L_p выступают общие свойства изображения: значение интенсивности (цветности) пикселя, тип текстуры, спектральные свойства изображения и т. п. Сюда относят методы порогового разделения, метод разрастания (увеличения) областей, метод деления областей. В методах второй группы реализуется отслеживание контуров и при определении L_p используется характеристика изменения интенсивности – разность величин и направления градиента, т. е. два пикселя (x, y) и (x', y') подобны, если имеет место соотношение:

$$L_p(S_i) = \left\| \left\| \nabla f(x, y) \right\| - \left\| \nabla f(x', y') \right\| \right\| \leq T, \\ \left| \varphi \nabla f(x, y) - \varphi \nabla f(x', y') \right| \leq A$$

где ∇ – градиент функции;
 φ – направление градиента;
 T, A – пороговые значения.

Исходными данными для задачи идентификации неиспользуемых земель являются многоспектральные снимки поверхности Земли, полученные при помощи космических спутников Sentinel-2, входящих в европейскую программу по мониторингу Земли «Copernicus» [8]. Спутники Sentinel-2 делают снимки в высоком разрешении: 10м для каналов в синем, зелёном, красном и ближнем инфракрасном диапазонах и 20м для большинства остальных каналов. Особенно важными для анализа растительности являются каналы, снимающие на красном крае фотосинтеза, и каналы красного и ближнего инфракрасного цветов, так как в этих диапазонах происходит резкое изменение в отражательной способности растения. Спутники Sentinel-2 работают в паре и обладают идентичным оборудованием, что позволяет получать обновлённые снимки поверхности с периодом в 5 дней.

В нашем случае, при задании L_p используются вегетационные индексы – показатели, рассчитываемые в результате операций с разными спектральными диапазонами (каналами) ДЗЗ, и имеющие отношение к параметрам растительности в данном пикселе снимка [9]. Существует множество вегетационных индексов, они в значительной степени похожи между собой. Они подбираются эмпирическим путём, исходя из известных особенностей кривых спектральной отражательной способности растительности и почв. Почти все распространённые вегетационные индексы используют только соотношение красного – ближнего инфракрасного каналов, предполагая, что в ближней инфракрасной области лежит линия открытой почвы. Подразумевается, что эта линия означает нулевое количество растительности [9].

Наиболее популярный и часто используемый вегетационный индекс NDVI (normalized difference vegetation index) рассчитывается на основе значений красного (R) и ближнего инфракрасного (NIR) диапазонов:

$$NDVI = \frac{(NIR - R)}{(NIR + R)}.$$

Значения индекса находятся в диапазоне от -1 до 1, т. е. «нормализованы» для более лёгкого сравнения между собой. Значения NDVI от -1 до 0 это объекты неживой природы и инфраструктуры – снег, вода, песок, камни, дома, дороги и т. п. Значения для растений лежат в диапазоне от 0 до 1.

Нейросетевой подход к идентификации.

Нейросетевой подход автоматизирует процесс идентификации, т. е. предикат L_p формируется неявно. Использование нейронных сетей обеспечивает выполнение главного требования к алгоритмам идентификации – их инвариантности к сдвигу объекта, его угловой ориентации и масштабу, что позволяет повысить достоверность идентификации.

При проведении исследования исходят из предположения, что совокупность сезонных значений NDVI залежей отличается от таковых у обрабатываемых земель. В связи с тем, что спектр залежи большую часть вегетационного периода слабо отличается от обычного поля, был выбран подход к получению серии снимков, сделанных на протяжении всего вегетационного периода. Будем считать, что спектральный отклик залежи на протяжении года более равномерный и стабильный, чем спектр пахотного поля. Хотя бы раз в год в спектре пашни могут быть обнаружены признаки недавней распашки. Также рабочая гипотеза предполагает, что текстура пахотных земель будет более равномерной за счет преобладания одной сельскохозяйственной культуры на всей площади поля, в то время как текстура неиспользуемых земель может быть неравномерна из-за различных травяных культур либо наличия кустарников. Сезонные значения NDVI, цветовые и текстурные информационные признаки могут быть использованы для обучения нейронной сети (НС) с целью повышения точности построения карт полей.

Нейросетевой алгоритм идентификации включает следующие шаги:

- выбор участка поверхности земли, содержащего как заброшенные, так и активно используемые сельскохозяйственные земли в примерно одинаковом соотношении;
- определение временных интервалов вегетационного периода, подбор спутниковых снимков выбранной территории сделанных на разных этапах развития растительности в течении одного сезона;
- расчёт и формирование карты NDVI для каждого отдельного снимка;
- совмещение опорных координат всех снимков и выделение общей области;
- формирование соответствующего этой области бинарного слоя данных, говорящего о принадлежности каждого пикселя залежным землям;
- обучение нейронной сети на тестовой выборке;
- оценка точности, построенной нейросети на валидационной выборке.

Для выбора территории (шаг 1), на которой будет произведена разметка данных, были использованы материалы геопортала земельно-информационной системы Республики Беларусь [10]. На рисунке 1 изображен участок территории Хойницкого района, часть территории которого активна распахана, а другая часть находится в зоне отчуждения после аварии на Чернобыльской АЭС и выведена из хозяйственной деятельности.

Отбор спутниковых снимков и их предварительная обработка (шаг 2) выполняются на сайте программы «Copernicus», где в параметрах поиска можно задать интересующую территорию, тип спутника, временной интервал, уровень облачности и др. Так как гипотеза идентификации основана на разнице отражательной способности залежей и обрабатываемой земли в течении вегетационного периода, то необходимо собрать как можно больше данных о развитии растительности на разных этапах, произвести отбор снимков, которые бы захватывали разные его этапы: посев, активная вегетация, различные даты сбора урожая для разных культур. Как показывают наши исследования, наиболее информативными являются снимки за осенний период, так как в это время спектральная характеристика используемых полей меняется сильнее чем спектр необрабатываемых территорий из-за уборки урожая и распашки территории. Последующие шаги алгоритма реализуются в программе QGIS (Open Source Geographic Information System) – свободной кроссплатформенной ГИС.

Технология обработки изображений в QGIS.

QGIS позволяет автоматически совмещать координаты полученных снимков, добавлять и редактировать растровые и векторные слои, рассчитывать вегетационные индексы на основании данных нескольких слоёв.

В результате из набора многоспектральных слоев будут получены слои значений NDVI, демонстрирующие состояние вегетации растительности в течении сезона (рисунок 2).



Рисунок 1. Изображение территории Хойницкого района

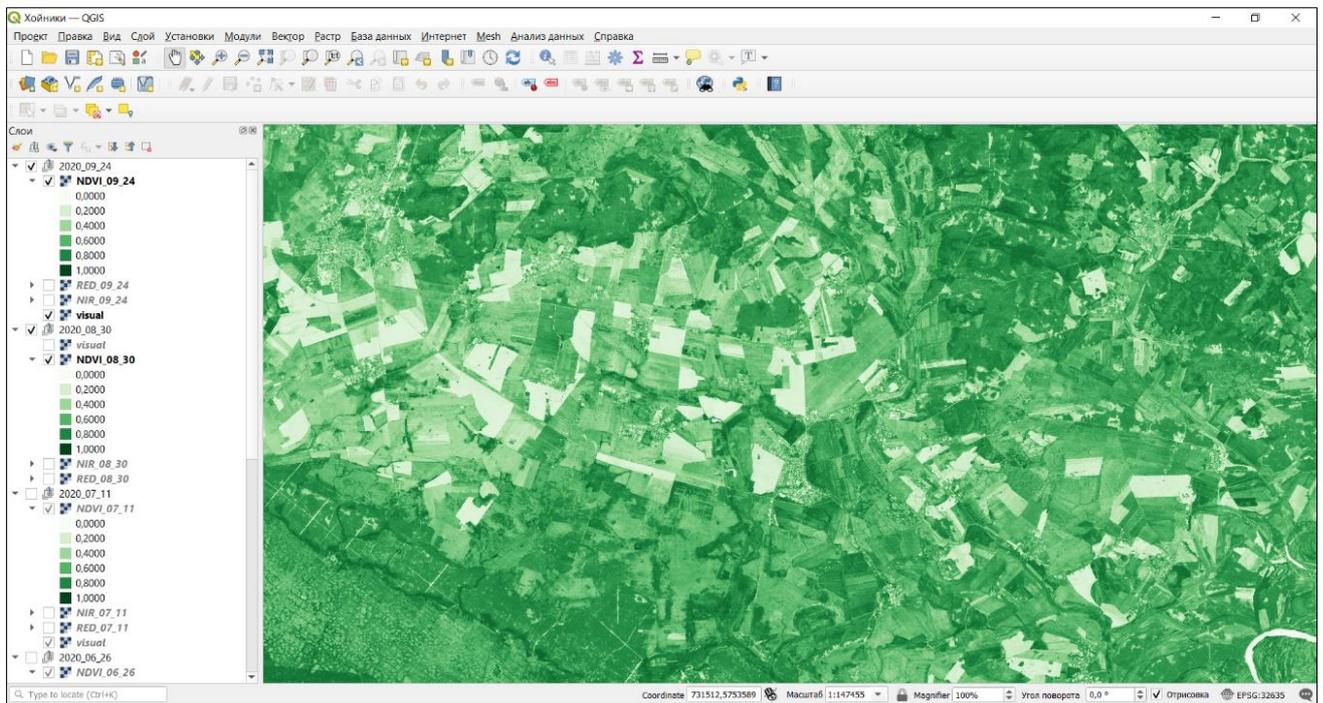


Рисунок 2. Визуализация карты значений NDVI

Далее для полученных данных сформируем обучающую выборку. Разметка данных производится на векторном слое путем добавления полигонов, разбитых на 2 класса: принадлежащих либо не принадлежащих залежным землям (рисунок 3). Построение полигонов намного менее трудозатратный процесс, чем задание класса каждому пикселю на растровом слое данных.

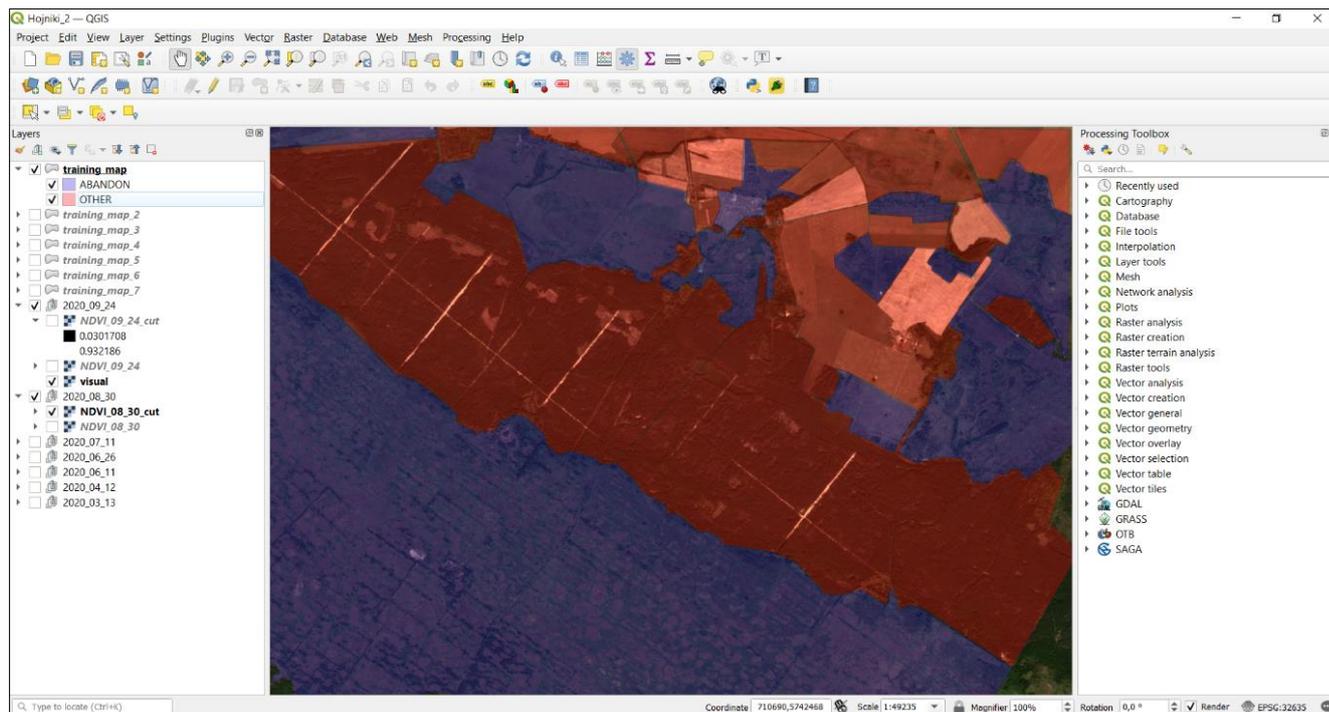


Рисунок 3. Заданная на векторном слое данных обучающая выборка

QGIS может быть адаптирована к особым потребностям благодаря расширяемой модульной архитектуре. QGIS предоставляет библиотеки, которые могут использоваться для создания модулей. Можно создавать отдельные приложения, используя языки программирования C++ или Python, которые разрабатываются сообществом. В настоящее время QGIS предоставляет возможность использовать инструменты анализа, выборки, геопроецирования, управления геометрией и базами данных.

В данной работе был использован модуль ОТВ (Orfeo ToolBox), открытый проект сообщества энтузиастов геоинформатики [11]. Модуль предоставляет множество решений по машинному обучению и BigData в ГИС: решает задачи сегментации, предоставляет методы машинного обучения, такие как различные типы классификации, искусственные нейронные сети, машины опорных векторов и др. На вход модуля обучения подается массив карт значений NDVI и векторный слой данных для обучения. В зависимости от метода обучения, есть возможность настраивать множество параметров сети (рисунок 4).

После обучения сети будет сформирован файл модели, который в последующем можно применить для идентификации залежных земель на других картах.

Заключение.

Предложен алгоритм идентификации залежных земель на основе расчета вегетационных признаков по многоспектральным изображениям ДЗЗ при использовании каналов красного и ближнего инфракрасного диапазона спектра. Алгоритм предусматривает сбор и использование данных вегетации полей на протяжении одного вегетационного сезона. Продемонстрированы возможности обработки изображений земной поверхности в программе QGIS и построения метода детектирования залежных земель с помощью модуля Orfeo Toolbox.

Работа выполнена при поддержке БРФФИ (белорусско-китайский проект № Ф20КИ-017 и белорусско-румынский проект № Ф20РА-014).

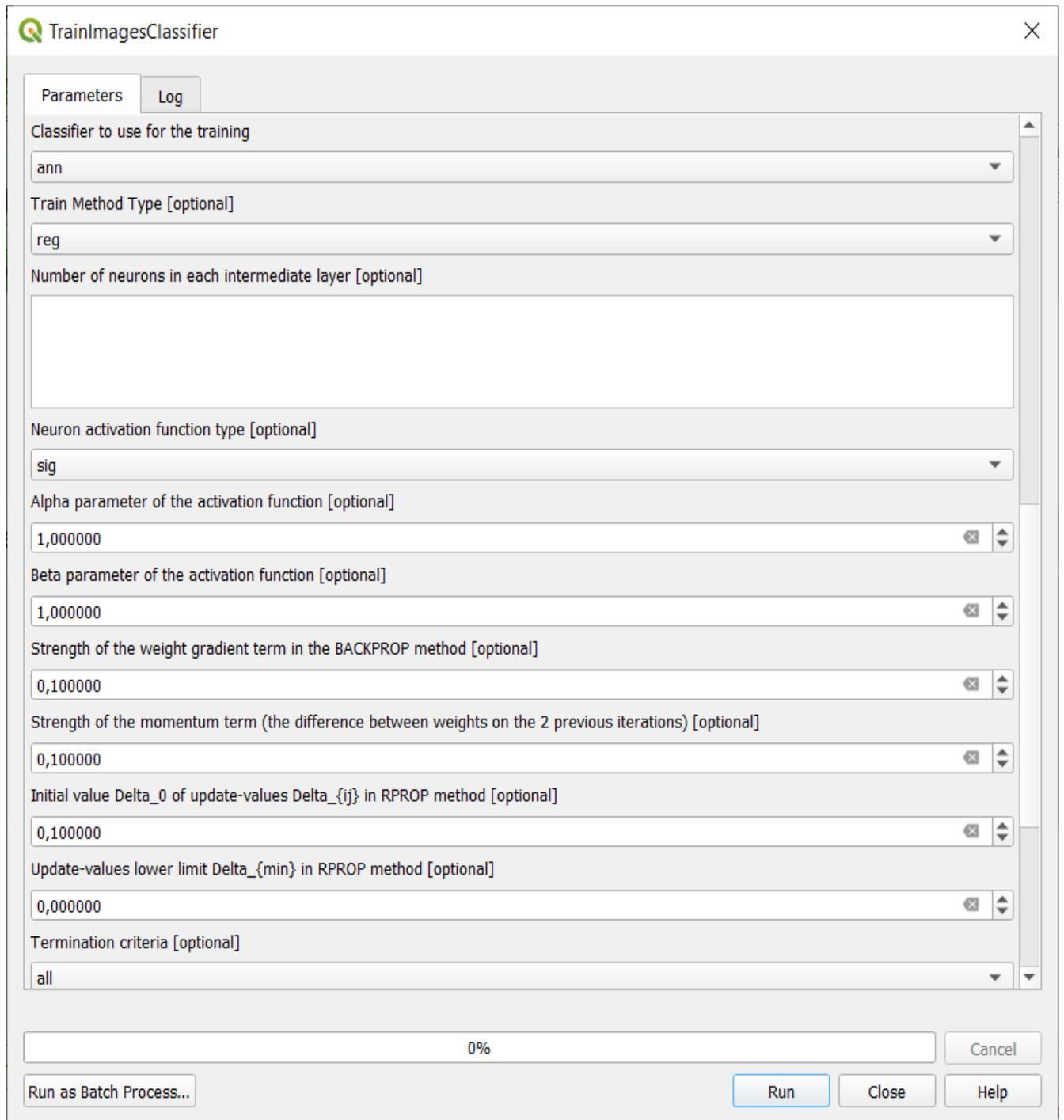


Рисунок 4. Настройка параметров сети ANN в Orfeo Toolbox

References

- [1] The effect of Landsat ETM/ETM+ image acquisition dates on the detection of agricultural land abandonment in Eastern Europe /Alexander V. Prishchepov [at al.] // Remote Sensing of Environment –2012. – Vol. 126. – Pp. 195-209.
- [2] Черепанов, А.С. Спектральные свойства растительности и вегетационные индексы / А.С. Черепанов, Е.Г. Дружинина. – Геоматика. – № 3. – 2009. – С. 28-32.
- [3] Терехин, Э.А. Распознавание залежных земель на основе сезонных значений вегетационного индекса NDVI / Э.А. Терехин // Компьютерная оптика. – 2017. – Т. 41, № 5. – С. 719-725.
- [4] Ganchenko, V. Agricultural Vegetation Monitoring Based on Aerial Data Using Convolutional Neural Networks / V. Ganchenko, A. Doudkin // Optical Memory and Neural Networks. – 2019. – Vol. 28, No. 2, pp. 129-134. © Allerton Press, Inc., 2019.

[5] Ганченко, В.В. Построение карт сельскохозяйственных полей по аэрофотоснимкам различного спектрального диапазона для систем точного земледелия / В.В. Ганченко, А.А. Дудкин, Е.Е. Марушко // Вестник БГТУ.-2017.-№ 5 (107).-С. 45-48.

[6] Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М.: Техносфера, 2005. – 1072 с.

[7] Введение в контурный анализ; приложения к обработке изображений и сигналов / Я.А. Фурман [и др.]. – М.: Физматлит, 2003. – 592 с.

[8] Copernicus Open Access Hub [сайт] URL: <https://scihub.copernicus.eu/dhus/#/home> (дата обращения: 16.03.2020).

[9] GIS LAB [сайт] URL: <https://gis-lab.info/qa/vi.html> (дата обращения: 28.02.2021).

[10] Геопортал земельно-информационной системы Республики Беларусь [сайт] URL: <https://gismap.by/next/> (дата обращения: 16.12.2020).

[11] Orfeo Toolbox [сайт] URL: <https://www.orfeo-toolbox.org/CookBook/index.html> (дата обращения: 16.02.2021).

NEURAL NETWORK IDENTIFICATION OF ABANDONED LANDS ON EARTH'S REMOTE SENSING IMAGES IN QGIS

A.M. LEVANKOU

Undergraduate student of the BSUIR, software engineer in LLC «Code Inspiration Pro»

A.A. DOUDKIN

Doctor of Computer Sciences, Professor, Head of the laboratory of System Identification of the United Institute of Informatics Problems the National Academy of Sciences of Belarus

*United Institute of Informatics Problems the National Academy of Sciences of Belarus
Belarusian State University of Informatics and Radioelectronics, Republic of Belarus
LLC «Code Inspiration Pro», Republic of Belarus
E-mail: alexlevankou@gmail.com, doudkin@newman.bas-net.by*

Abstract. This work solves the problem of identifying unused and fallow (abandoned) lands on Earth's remote sensing images. The initial data are multispectral images obtained using the Sentinel-2 space satellites, and a set of vegetation indices, based on which the state of vegetation is assessed. A training data set is created, based on which a neural network identification algorithm is built. The result of identification is a map of unused lands. Data processing is carried out based on the open geographical information system QGIS.

Keywords: Earth's remote sensing, multispectral image, abandoned agricultural land, fallow land, vegetation index, identification, artificial neural network, geographical information system, GIS, NDVI, QGIS, Orfeo Toolbox.