

## Применение методов машинного обучения для поиска ковулканических ионосферных возмущений по данным ГНСС-наблюдений

А. С. Тен<sup>1</sup>, Н. В. Шестаков<sup>2,6</sup>, А. А. Сорокин<sup>1</sup>,  
Н. Н. Титков<sup>3</sup>, М. Озоно<sup>4</sup>, Х. Такахаша<sup>5</sup>

<sup>1</sup> *Вычислительный центр ДВО РАН, Хабаровск, 680000, Россия*  
*E-mail: alexander.s.ten@yandex.ru*

<sup>2</sup> *Институт прикладной математики ДВО РАН, Владивосток, 690041, Россия*

<sup>3</sup> *Камчатский филиал ФИЦ ЕГС РАН*  
*Петропавловск-Камчатский, 683023, Россия*  
*E-mail: nik@emsd.ru*

<sup>4</sup> *Институт исследования землетрясений, Токийский университет*  
*Токио, Япония*  
*E-mail: m.ohzono@sci.hokudai.ac.jp*

<sup>5</sup> *Институт сейсмологии и вулканологии, Университет Хоккайдо,*  
*Саппоро, Япония*  
*E-mail: hiroaki@sci.hokudai.ac.jp*

<sup>6</sup> *Дальневосточный федеральный университет, Владивосток, 690092, Россия*  
*E-mail: shestakov.nv@dyfu.ru*

Работа посвящена оценке возможности применения искусственных нейронных сетей для поиска ковулканических ионосферных возмущений во временных рядах полного электронного содержания, полученных по данным ГНСС-наблюдений. На примере извержения влк. Пик Сарычева 11–16 июня 2009 г. определены и размечены инструментальные ГНСС-данные, сгенерированы их наборы с разными размерами образцов и двумя классами: с возмущениями и без них. На основе этой информации обучены и исследованы пять искусственных нейронных сетей различных архитектур, применяемых для решения задачи классификации временных рядов. Вычислены метрики качества классификации и проведено их сравнение. Получены результаты тестирования предложенного алгоритма с разными классификаторами — нейронными сетями, показавшими лучший результат: InceptionTime и ResNet. Проведён анализ и сравнение этих результатов между собой и с результатами алгоритма STA/LTA по количеству найденных ковулканических возмущений, ложных срабатываний и скорости работы на тестовых файлах. Описаны недостатки STA/LTA относительно применения к исследуемым данным и возможности их преодоления в предложенном подходе. Сделаны выводы о качестве классификации нейронных сетей и их применимости как классификатора в универсальном алгоритме поиска ковулканических возмущений. Предложены направления будущих исследований по рассматриваемой теме.

**Ключевые слова:** ионосфера, ковулканические возмущения, машинное обучение, искусственные нейронные сети, дистанционное зондирование, ГНСС

Одобрена к печати: 19.01.2023  
DOI: 10.21046/2070-7401-2023-20-1-37-54

### Введение

Использование данных о полном электронном содержании (ПЭС) в ионосфере Земли, получаемых на основе наблюдений в ГНСС-сетях (ГНСС — глобальные навигационные спутниковые системы), для исследования таких опасных природных явлений, как землетрясения (Aji et al., 2017; Lin, Chiou, 2020), цунами (Galvan et al., 2011), геомагнитные бури (Astafyeva et al., 2015), рассматривается всё большим количеством исследователей. В последние годы большое значение приобретают работы, связанные с использованием данных ГНСС-сетей в исследованиях ответной реакции ионосферы Земли на извержения вулканов, особенно в связи с мощнейшим подводным извержением 15 января 2022 г. влк. Хунга-Тонга-Хунга-

Хаапай, вызвавшим разрушительное цунами и многочисленные эффекты в атмосфере Земли (Muafiy et al., 2022). В них реконструируются значения ПЭС, представленные в виде временного ряда, которые затем исследуются на наличие ковулканических ионосферных возмущений (КИВ). Например, в работе (Shestakov et al., 2021) изучаются КИВ, вызванные извержением влк. Пик Сарычева 11–16 июня 2009 г. Есть и другие подобные работы (Неки, 2006; Shults et al., 2016).

Большой объём инструментальной информации усложняет её обработку специалистом-оператором, что значительно увеличивает время на анализ события и риск возникновения различных технических ошибок. В целом эти и другие ограничения пока не позволяют в полной мере исследовать КИВ, а также осуществлять оперативный мониторинг вулканов с использованием данных ПЭС. Одним из подходов к решению указанной проблемы представляется применение методов машинного обучения — искусственных нейронных сетей — для автоматизированного обнаружения КИВ в данных ПЭС. Актуальность исследования обусловлена отсутствием эффективных компьютерных решений для детектирования КИВ в рядах ПЭС, полученных по данным ГНСС-измерений.

Цель настоящего исследования — рассмотреть применимость некоторых современных архитектур нейронных сетей для анализа инструментальных данных ГНСС-измерений с целью выявления и классификации ковулканических возмущений.

### Исходные инструментальные данные для поиска КИВ

Объектом исследования выбран стратовулкан Пик Сарычева (высота 1446 м над уровнем моря), расположенный на о. Матуа, который относится к Большой Курильской гряде и находится на расстоянии 400 км к юго-западу от оконечности п-ова Камчатка (Российская Федерация) и 750 км к северо-востоку от о. Хоккайдо (Япония) (рис. 1). Сильное эксплозивное извержение влк. Пик Сарычева (индекс эксплозивности VEI (англ. Volcanic Explosivity Index) равен 4) произошло 11–16 июня 2009 г. после 33-летнего периода покоя.

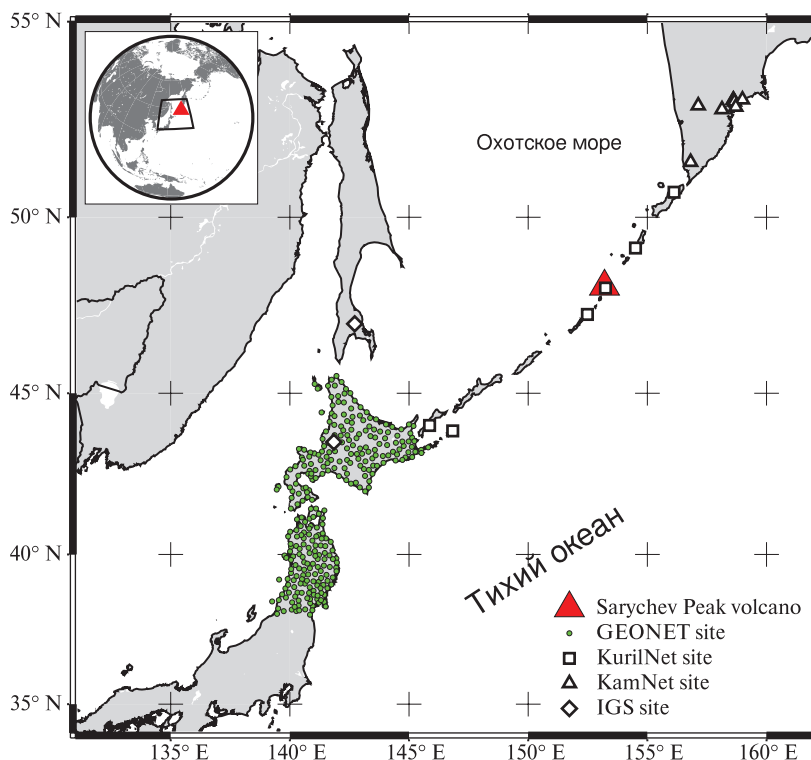
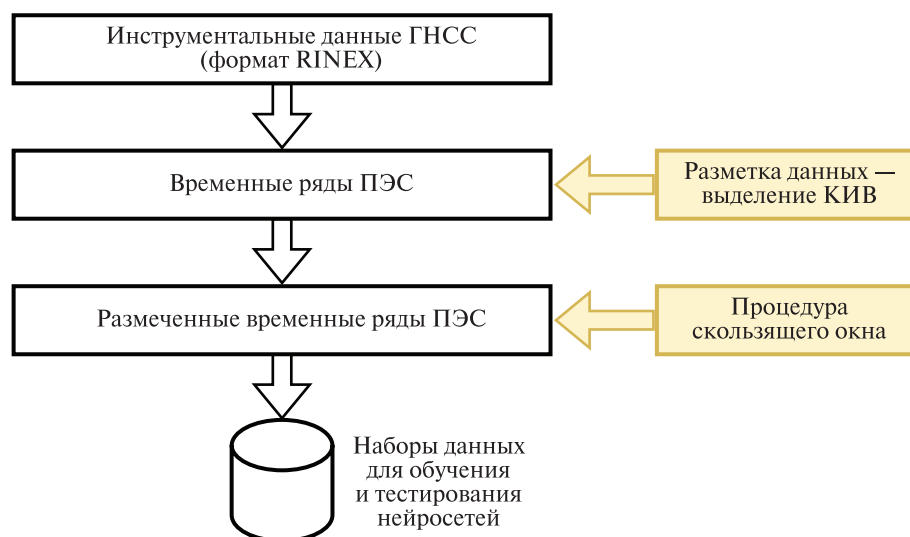


Рис. 1. Район расположения влк. Пик Сарычева с информацией о местах установки ГНСС-станций. Рисунок заимствован из статьи (Shestakov et al., 2021)

В качестве экспериментальных данных в рамках решения поставленных задач был использован архив с 18 958 записями временных рядов ПЭС за 15 и 16 июня 2009 г., которые реконструированы из файлов исходных спутниковых наблюдений на станциях преимущественно японской ГНСС-сети GEONET (северная её часть, см. *рис. 1*) (Tsuji et al., 2017), записанных в формате RINEX и содержащих измеренные псевдодальности и набег фазы спутников GPS (*англ.* Global Positioning System — система глобального позиционирования) на частотах  $L1 = 1575,42$  МГц и  $L2 = 1227,60$  МГц.

В общей сложности в работе использованы данные 318 станций различных ГНСС-сетей: японской сети GEONET (Tsuji et al., 2017) — 302 станции; глобальной сети IGS (*англ.* International GNSS Service, <http://www.igs.org/>) — 2 станции; Курильской ГНСС-сети KurilNet (Kogan et al., 2013) — 6 станций; Камчатской ГНСС-сети KamNet (Levin et al., 2014) — 8 станций. Следует отметить, что пункты сети GEONET расположены на расстоянии 20–30 км друг от друга, что обеспечивает высокую плотность покрытия треками ионосферных точек области вокруг вулкана. Данные сети KurilNet были получены нами на сайте UNAVCO (*англ.* University NAVSTAR Consortium) (<https://www.unavco.org/data/gps-gnss/gps-gnss.html>).

На *рис. 2* представлен алгоритм формирования наборов данных для обучения и тестирования нейронных сетей.



*Рис. 2.* Алгоритм подготовки данных для последующего обучения нейронных сетей

На первом этапе с помощью утилиты TEC-SUITE (<http://www.gnss-lab.org/tec-suite.html>) реконструировались значения ПЭС из исходных файлов, записанных в формате RINEX и содержащих суточные сеансы наблюдений. Далее к полученному результату применялся фильтр скользящего среднего. Полученные значения сохранялись в текстовых файлах, соответствующих парам «станция – спутник». Шаг дискретизации записей — 30 с. Особенностью полученных временных рядов стал тот факт, что они могут содержать пробелы (*рис. 3*, см. с. 40), соответствующие периодам невидимости космического аппарата и обусловленные сбоями в работе спутниковой аппаратуры и другими причинами, например такими, как солнечная активность, геомагнитные штормы, глубокое ультрафиолетовое излучение и прочие факторы (Zewdie et al., 2021).

На следующем этапе в ручном режиме проводилась разметка полученного архива ПЭС: по данным из работы (Shestakov et al., 2021) выделялись сегменты с возмущениями, которые оказались следствием эксплозивных эпизодов извержения влк. Пик Сарычева за 14 и 15 июня 2009 г. КИВ представляют собой волновые формы характерного вида различной продолжительности и природы. На *рис. 4а, б* (см. с. 40) представлены примеры типичных волновых форм и размеченных КИВ.

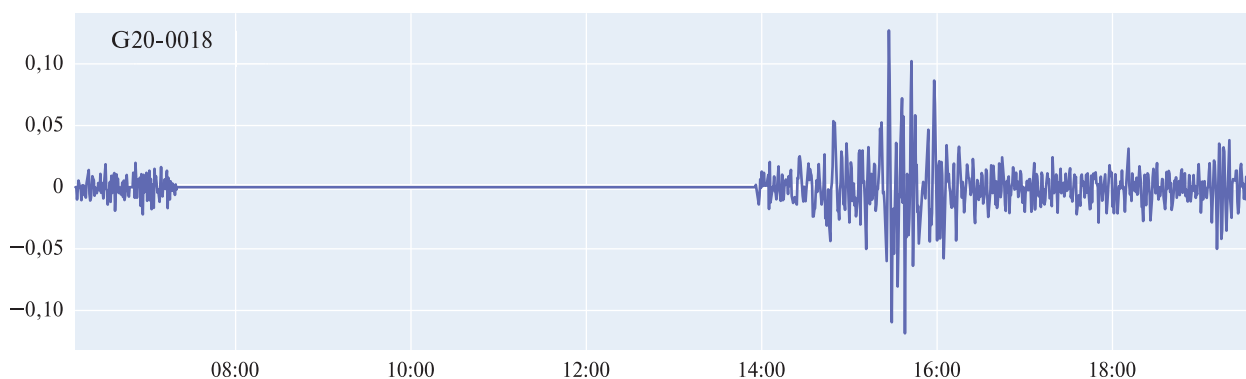
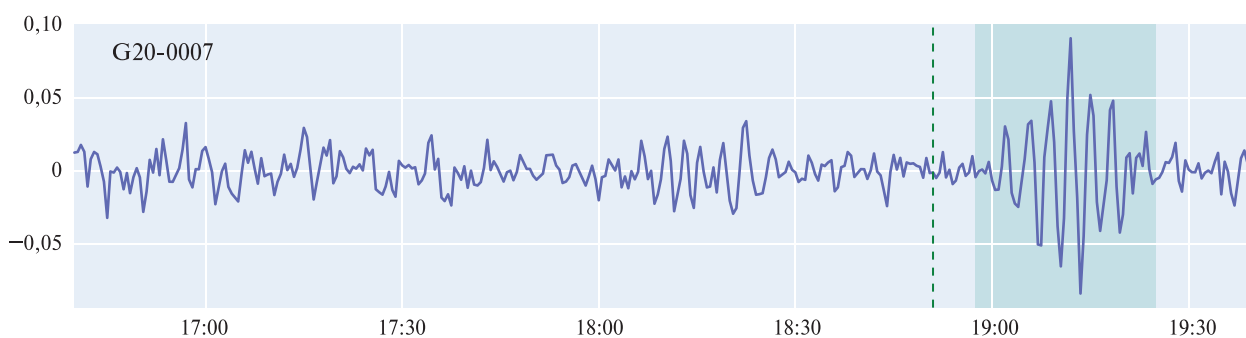
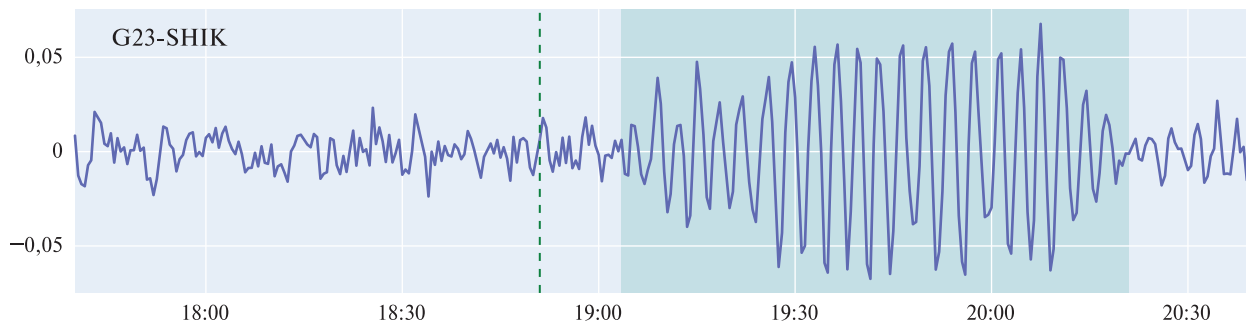


Рис. 3. Пример графика временного ряда ПЭС — волновой пакет от ст. 0018 сети GEONET и спутника G20 (GPS) за 14 июня 2009 г. Длительный пробел в данных в период времени 07:19–13:55 UT (англ. Universal Time ) изображён прямой линией



*a*



*б*

Рис. 4. Графики временных рядов ПЭС с размеченными ковулканическими возмущениями (выделены зелёным цветом), являющимися следствием одного из эксплозивных эпизодов извержения влк. Пик Сарычева, произошедшего 14 июня 2009 г.: *a* — волновой пакет от ст. 0007 сети GEONET и спутника G20 (GPS); *б* — волновой пакет от ст. SHIK сети KurilNet и спутника G23 (GPS). На графиках показаны разные типы волновых пакетов, инициированных извержением. В левом верхнем углу указаны тип и системный номер спутника (G — GPS) и аббревиатура/код пункта наблюдений. Вертикальной пунктирной линией дано время эруптивного события

На завершающем этапе из размеченных данных генерировались выборки для обучения нейронных сетей: для каждого размеченного файла с помощью процедуры скользящего окна извлекались подпоследовательности заданного размера (окна) с заданным шагом, содержащие либо часть КИВ, либо фоновый шум. Таким образом, были получены пять наборов данных с образцами разной длины: 5, 10, 20, 30 и 50 значений ПЭС. Каждая выборка содержит представленные поровну образцы двух классов: 1) нет возмущений (шум), 2) есть возмущения.

Перед обучением нейронных сетей каждый набор данных проходил процедуру нормализации и стандартизации (Ahsan et al., 2021). Эксперименты с обучением нейронных сетей на полученных наборах данных показывают, что такая их предобработка существенно улучшает сходимость обучения.

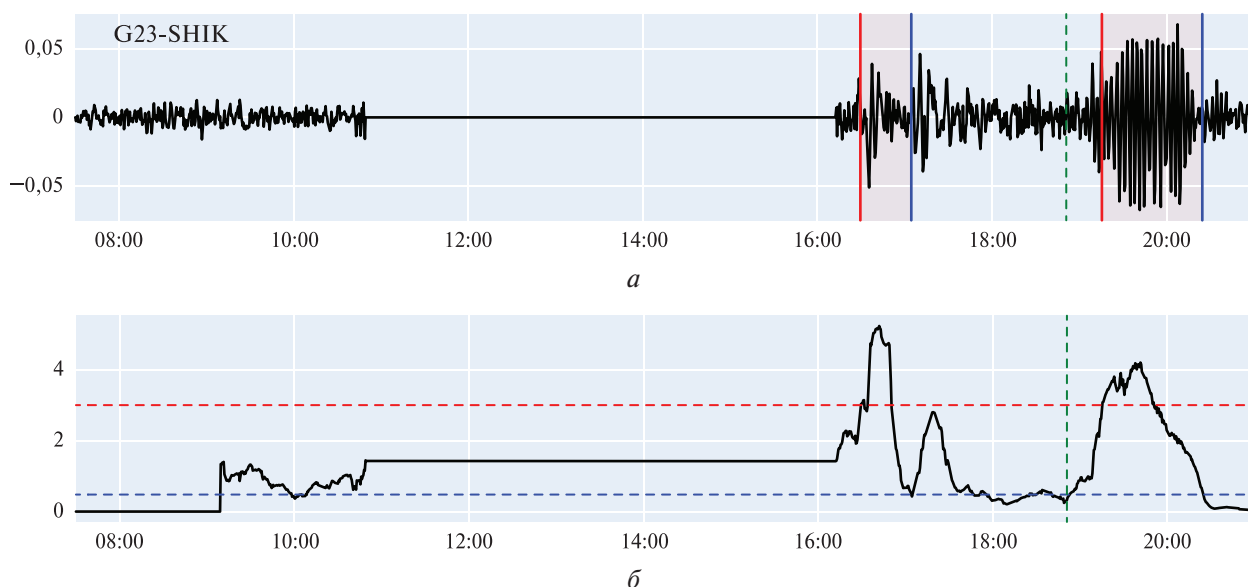
### Алгоритм STA/LTA для поиска КИВ

В целях сравнения с предложенным подходом в качестве базового был выбран алгоритм STA/LTA, широко использующийся в сейсмологии для автоматического обнаружения сейсмических событий на сейсмограммах (Han et al., 2010; Withers et al., 1998).

Алгоритм вычисляет отношение (STA/LTA) плотности энергии сигнала в коротком временном окне (*англ.* short time average — STA) и в длинном временном окне (*англ.* long time average — LTA). В результате получается новый ряд значений, соответствующих исходным значениям входного ряда. Далее для полученного ряда значений превышение верхнего порогового значения выступает триггером начала искомой волновой формы, а достижение нижнего порога — триггером конца. Алгоритм имеет четыре основных параметра, подбираемых эмпирически: размеры короткого и длинного окон и два значения порогов (триггеров) — срабатывания (верхний) и отключения (нижний) ([https://gfzpublic.gfz-potsdam.de/pubman/item/item\\_4097](https://gfzpublic.gfz-potsdam.de/pubman/item/item_4097)).

Проведены эксперименты по применению алгоритма для обнаружения КИВ в рядах ПЭС. Использовалась реализация из библиотеки ObsPy (Krischer et al., 2015). Пороговые значения триггеров срабатывания и отключения вычислялись по формулам:  $\text{threshold}_{on} = \mu + k_1\sigma$ ,  $\text{threshold}_{off} = \mu - k_2\sigma$ , где  $\mu$  — математическое ожидание значений STA/LTA;  $\sigma$  — стандартное отклонение;  $k_1, k_2$  — задаваемые коэффициенты. В ходе экспериментов были подобраны следующие значения параметров: размер окна STA — 30, LTA — 200 (количество временных шагов), коэффициенты  $k_1 = 1,5$  и  $k_2 = 0,5$ . Пробелы в данных игнорировались.

Пример детектирования КИВ приведён на *рис. 5*. Алгоритм выделил два участка ряда: первый (ранний по времени на *рис. 5*) — не является КИВ (ложное срабатывание), а второй — реальное КИВ.



*Рис. 5.* Временной ряд ПЭС (а) (волновой пакет от ст. SHIK сети KurilNet и спутника G23 (GPS) за 14 июня 2009 г.); вертикальные красная и синяя линии — начало и конец возмущения, определённые алгоритмом. Соответствующее значение STA/LTA (б); красная и синяя пунктирные линии — пороговые уровни триггеров срабатывания и отключения. Вертикальной зелёной пунктирной линией выделено время эруптивного события

Эксперименты со сканированием 1000 размеченных файлов данных ПЭС показывают следующий результат: обнаружено 26 из 49 КИВ и 1163 ложных срабатываний. При этом в алгоритм были добавлены дополнительные условия на продолжительность найденных возмущений — не более двух часов.

К основным недостатком алгоритма применительно к исследуемым данным ПЭС можно отнести его неспособность различать возмущения различной природы, т.е. любое возмущение достаточной большой амплитуды будет детектировано. Также особенность данных ПЭС — наличие пробелов — не позволяет полноценно применять рассматриваемый алгоритм, так как для вычисления текущих значений STA/LTA необходима информация о предыдущих значениях ряда, которая может в таких данных отсутствовать.

## Классификация временных рядов ПЭС методами машинного обучения

Другой подход, применённый нами, состоит в тестировании нейронных сетей нескольких архитектур, обученных на полученных наборах данных, и их сравнении по метрикам качества классификации, а также в анализе работы алгоритма детектирования КИВ на основе процедуры скользящего окна и обученной нейронной сети, используемой в качестве классификатора на тестовых файлах, выделенных из имеющихся экспериментальных данных.

Для решения поставленной задачи был рассмотрен алгоритм на основе процедуры скользящего окна и классификатора, который должен на вход принимать файлы ПЭС, а на выходе выдавать время начала и конца найденных сегментов, классифицированных как класс «с возмущением». Определяющая часть алгоритма — классификатор, в качестве которого выступает нейронная сеть, обученная классифицировать входящие сегменты временного ряда, содержащие или не содержащие искомые ковулканические возмущения. Подобный подход рассматривался в работе (Brissaud, Astafyeva, 2022) для поиска косейсмических ионосферных возмущений по данным ГНСС, однако в качестве классификатора использовалась не нейронная сеть, а метод машинного обучения Random Forest.

В теории предложенный нами подход лишён вышеуказанных недостатков применения STA/LTA — нейронные сети могут быть обучены разделять волновые формы разной природы (например, ковулканические возмущения и возмущения, вызванные солнечной активностью). Также важным достоинством такого подхода становится отсутствие требования непрерывности анализируемого временного ряда, так как каждый отдельный образец, извлечённый процедурой скользящего окна, обрабатывается независимо.

На *рис. 6* (см. с. 43) изображена блок-схема работы предложенного алгоритма. Из исходного файла с временным рядом ПЭС с помощью процедуры скользящего окна получается сегмент  $w^t$ , который затем классифицируется. Если сегмент распознан как содержащий возмущение, то время его начала и конца сохраняется в хранилище и происходит переход к следующему сегменту  $w^{t+1}$ . После обработки всего архива при необходимости проводится постобработка (если данные предварительно размечены) — вычисляется количество найденных размеченных КИВ, а также количество ложных срабатываний. КИВ считается найденным, если хотя бы один сегмент, классифицированный как сегмент «с возмущением», пересекается с размеченным сегментом КИВ на заданное минимальное количество временных шагов (в данной работе оно равнялось пяти временным шагам).

Вопрос выбора подходящего метода для классификатора представляется нетривиальным. Методы классификации временных рядов включают в себя обширный список успешно применяемых в различных областях как классических методов машинного обучения (без использования глубокого обучения), таких как Random Forest (Nguyen et al., 2020), методы на основе DTW (Li et al., 2020), методы с применением вейвлет-преобразования (Azar et al., 2020), NIVE-COTE v2 (Middlehurst et al., 2021), так и подходы с применением глубокого обучения (Fawaz et al., 2019).

Преимущество большинства классических методов — их интерпретируемость, т.е. степень того, насколько человек-эксперт может понять причину принятия решения моделью

машинного обучения. Однако к их недостаткам можно отнести большую вычислительную сложность, например сложность NIVE-COTE v2 составляет  $O(n^2l^2)$ , где  $n$  — размер обучающей выборки,  $l$  — длина образца (временного ряда). Вывод (англ. inference) обученной нейронной сети работает очень быстро, но само обучение становится длительным и сложным процессом. Таким образом, приоритет скорости работы классификатора в алгоритме детектирования КИВ указывает на целесообразность использования обученной нейронной сети.

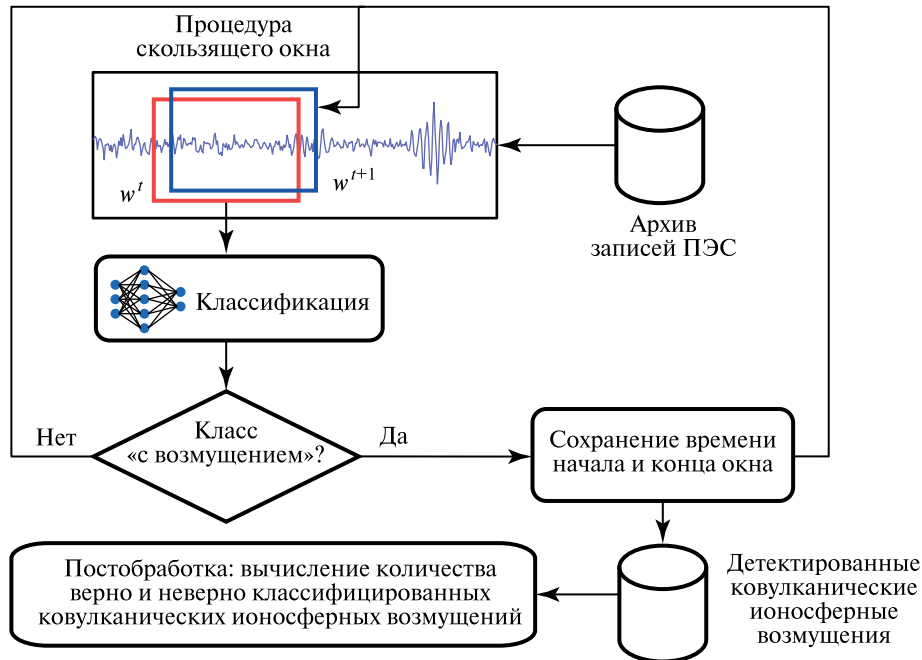


Рис. 6. Блок-схема работы алгоритма поиска КИВ

Таблица 1. Рассматриваемые архитектуры нейронных сетей

Архитектура	Слои для извлечения признаков	Функция активации	Регуляризация/нормализация	Примечание
MLP (Wang et al., 2017)	Полносвязные слои (англ. Fully connected)	ReLU/Leaky ReLU	Dropout/–	Архитектура из простых полносвязных слоёв
FCN (Usmankhujaev et al., 2022)	Слой глобального пулинга (англ. Global Average Pooling)	ReLU	–/BatchNorm	Преобладание свёрточных слоёв
ResNet (Wang et al., 2017)	Слой MaxPooling			Механизм пропуска слоёв
InceptionTime (Fawaz et al., 2020)	Кодирующие и декодирующие слои	GELU	Dropout/ BatchNorm	Ансамбль пяти сетей Inception
Time Series Transformer (Zerveas et al., 2021)				Механизм внимания

Примечание: MLP — англ. multilayer perceptrons, FCN — англ. fully convolutional network.

Определяющими факторами при выборе нейронной архитектуры в нашем исследовании стали:

- скорость обучения и скорость вывода нейронной сети (которые зависят как от количества параметров нейронной сети, так и от типов применяемых слоёв);
- качество классификации временных рядов различной длины. Указанный выбор обусловлен необходимостью в перспективе обрабатывать большие массивы данных в режиме реального времени.

Также важно, чтобы нейронная сеть в процессе обучения эффективно извлекала необходимые признаки волновых форм КИВ и шума для точной их классификации, т. е. обеспечивала высокую обобщающую способность.

В соответствии с приведёнными условиями были отобраны пять архитектур нейронных сетей (*табл. 1*).

Дополнительно для сравнения в работе приводятся значения метрик качества классификации методов Random Forest и NIVE-COTE v2. Random Forest — универсальный алгоритм машинного обучения, заключающийся в использовании ансамбля решающих деревьев (Breiman, 2001; Nguyen et al., 2020). NIVE-COTE версии 2.0 — передовой алгоритм для классификации временных рядов, он представляет собой мета-ансамбль из четырёх разных классификаторов (Middlehurst et al., 2021).

В работе используются метрики оценки классификации accuracy, precision, recall и  $F_1$ -мера ( $F_1$ -score), вычисленные для каждого класса в выборке:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \text{precision}_{\text{КИВ}} = \frac{TP}{TP + FP}, \quad \text{recall}_{\text{КИВ}} = \frac{TP}{TP + FN},$$

$$\text{precision}_{\text{шум}} = \frac{TN}{TN + FN}, \quad \text{recall}_{\text{шум}} = \frac{TN}{TN + FP}, \quad F_1 = 2 \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}} = \frac{TP}{TP + (FP + FN)/2},$$

где TP (*англ.* True Positive) — количество примеров, верно классифицированных как принадлежащие рассматриваемому классу; TN (*англ.* True Negative) — количество примеров, верно классифицированных как не принадлежащие рассматриваемому классу; FP (*англ.* False Positive) — количество примеров, неверно классифицированных как принадлежащие рассматриваемому классу; FN (*англ.* False Negative) — количество примеров, неверно классифицированных как не принадлежащие рассматриваемому классу. Положительные примеры — это примеры с КИВ, отрицательные — с шумом.

Accuracy понимается как доля примеров, для которых классификатор правильно определил их класс: шум или возмущение; precision — доля примеров, верно отнесённых классификатором к положительному (отрицательному) классу среди всех примеров, отнесённых к положительному (отрицательному) классу; recall — доля примеров, верно отнесённых классификатором к положительному (отрицательному) классу среди всех примеров, отнесённых к положительному (отрицательному) классу. Например, относительно класса с КИВ precision показывает, сколько примеров из определённых нейронной сетью как КИВ действительно являются примерами с КИВ, а recall — какую долю примеров с КИВ нейронной сети удалось выявить. Аналогично и для класса шума.  $F_1$ -мера — это скомпонованные в одну метрику precision и recall.

## Полученные результаты и их обсуждение

На наборах данных были обучены и протестированы нейронные сети пяти архитектур, наиболее подходящих под требования настоящего исследования. В работе использованы реализации: нейронных сетей из библиотеки tsai (<https://github.com/timeseriesAI/tsai>), алгоритмов Random Forest и NIVE-COTE v2 из библиотек pyts (Faouzi, Janati, 2020) и sktime (Löning et al., 2019) соответственно.

При обучении применялась k-fold кросс-валидация — процедура эмпирического оценивания обобщающей способности нейронных сетей. Параметр  $k$  принят равным пяти — пять экземпляров каждой из рассматриваемых архитектур нейронных сетей обучались и тестировались с разным разделением обучающей и тестовой выборок. В *табл. 2–8* приведены усреднённые значения метрик оценки классификации со стандартным отклонением для каждой из архитектур и набора данных (полужирным начертанием выделены лучшие значения). Каждая выборка делилась на обучающую часть (80 %) и тестовую (20 %). Количество эпох обучения — 100.

Информация о количестве параметров каждой нейронной сети (для каждого набора данных по размеру образца) приведена в *табл. 9*. Там же дано время вывода каждой нейронной



сети. Указаны среднее значение и стандартное отклонение измеренного времени вывода с повторением в 300 итераций, единицы измерения — миллисекунды. Измерения времени вывода производились на одном графическом ускорителе Nvidia A100.

Метрики классификации различаются в зависимости от размера окна: для обучающей выборки с окном большего размера значения всех трёх метрик классификации выше. Большой размер окна более пригоден для сканирования архивов ПЭС, однако для использования в алгоритмах реального времени может понадобиться окно меньшего размера. Нейронные сети архитектур InceptionTime и ResNet продемонстрировали самые высокие показатели метрик по всем размерам окна. InceptionTime имеет наименьшее количество параметров почти везде, количество параметров напрямую влияет на время обучения нейронной сети. Указанные нейронные сети почти во всех случаях превзошли по значениям метрик алгоритмы Random Forest и HIVE-COTE v2 либо показали очень близкие значения.

Таблица 2. Результаты тестирования обученных нейронных сетей и алгоритмов — значения accuracy

Нейронная сеть (алгоритм)	Размер образца				
	5	10	20	30	50
MLP	0,82±0,01	0,65±0,22	0,91±0,01	0,91±0,01	0,94±0,02
FCN	0,84±0,02	<b>0,90±0,03</b>	<b>0,96±0,01</b>	0,96±0,01	0,98±0,01
ResNet	<b>0,85±0,02</b>	0,88±0,02	<b>0,96±0,01</b>	<b>0,97±0,01</b>	<b>0,98±0,00</b>
InceptionTime	0,84±0,02	<b>0,90±0,03</b>	<b>0,96±0,01</b>	0,96±0,01	0,98±0,01
TSTransformer	0,82±0,01	0,78±0,2	0,86±0,02	0,86±0,01	0,86±0,04
Random Forest	0,81±0,02	0,88±0,01	0,91±0,02	0,90±0,01	0,92±0,02
HIVE-COTE v2	0,69±0,02	0,83±0,01	0,92±0,02	0,93±0,02	0,98±0,01

Таблица 3. Результаты тестирования обученных нейронных сетей и алгоритмов — значения precision относительно класса «шум»

Нейронная сеть (алгоритм)	Размер образца				
	5	10	20	30	50
MLP	0,78±0,02	0,81±0,28	0,89±0,02	0,89±0,02	0,91±0,03
FCN	0,84±0,02	<b>0,90±0,06</b>	0,95±0,01	0,96±0,02	0,97±0,01
ResNet	<b>0,85±0,03</b>	0,85±0,05	<b>0,96±0,01</b>	<b>0,97±0,02</b>	<b>0,98±0,01</b>
InceptionTime	0,83±0,03	0,86±0,1	0,96±0,02	0,96±0,02	<b>0,98±0,01</b>
TSTransformer	0,78±0,02	0,86±0,08	0,83±0,03	0,83±0,03	0,82±0,06
Random Forest	0,80±0,03	0,88±0,01	0,90±0,03	0,89±0,01	0,91±0,03
HIVE-COTE v2	0,70±0,02	0,86±0,02	0,96±0,02	<b>0,97±0,02</b>	<b>0,99±0,01</b>

Таблица 4. Результаты тестирования обученных нейронных сетей и алгоритмов — значения precision относительно класса «с возмущением»

Нейронная сеть (алгоритм)	Размер образца				
	5	10	20	30	50
MLP	<b>0,88±0,01</b>	0,66±0,23	0,94±0,02	0,94±0,01	0,97±0,01
FCN	0,84±0,03	0,91±0,02	0,96±0,01	<b>0,97±0,01</b>	<b>0,98±0,01</b>
ResNet	0,85±0,02	0,92±0,03	<b>0,97±0,01</b>	<b>0,97±0,01</b>	<b>0,98±0,01</b>
InceptionTime	0,85±0,02	<b>0,94±0,02</b>	<b>0,97±0,01</b>	<b>0,97±0,01</b>	<b>0,98±0,01</b>
TSTransformer	0,87±0,01	0,82±0,2	0,92±0,02	0,90±0,02	0,92±0,03
Random Forest	0,83±0,02	0,89±0,02	0,92±0,02	0,91±0,01	0,92±0,02
HIVE-COTE v2	0,68±0,02	0,80±0,01	0,89±0,03	0,91±0,05	0,96±0,02

Таблица 5. Результаты тестирования обученных нейронных сетей и алгоритмов — значения recall относительно класса «шум»

Нейронная сеть (алгоритм)	Размер образца				
	5	10	20	30	50
MLP	<b>0,89±0,01</b>	0,38±0,49	0,95±0,02	0,94±0,01	0,97±0,01
FCN	0,84±0,04	0,91±0,03	0,96±0,01	<b>0,97±0,01</b>	<b>0,98±0,01</b>
ResNet	0,85±0,02	0,92±0,04	<b>0,97±0,01</b>	<b>0,97±0,01</b>	<b>0,98±0,01</b>
InceptionTime	0,85±0,02	<b>0,95±0,02</b>	<b>0,97±0,01</b>	<b>0,97±0,01</b>	<b>0,98±0,01</b>
TSTransformer	<b>0,89±0,01</b>	0,73±0,4	0,93±0,02	0,91±0,03	0,93±0,03
Random Forest	0,84±0,03	0,89±0,02	0,92±0,02	0,91±0,01	0,92±0,02
HIVE-COTE v2	0,66±0,02	0,78±0,01	0,88±0,04	0,90±0,05	0,96±0,02

Таблица 6. Результаты тестирования обученных нейронных сетей и алгоритмов — значения recall относительно класса «с возмущением»

Нейронная сеть (алгоритм)	Размер образца				
	5	10	20	30	50
MLP	0,75±0,03	<b>0,91±0,08</b>	0,88±0,02	0,88±0,03	0,91±0,04
FCN	<b>0,84±0,03</b>	0,90±0,07	0,95±0,01	0,96±0,02	0,97±0,01
ResNet	<b>0,84±0,03</b>	0,83±0,07	<b>0,96±0,01</b>	<b>0,97±0,02</b>	<b>0,98±0,01</b>
InceptionTime	0,83±0,03	0,84±0,1	0,96±0,02	0,96±0,02	<b>0,98±0,01</b>
TSTransformer	0,75±0,03	0,84±0,1	0,80±0,05	0,81±0,04	0,79±0,08
Random Forest	0,79±0,03	0,87±0,01	0,89±0,03	0,88±0,01	0,92±0,03
HIVE-COTE v2	0,72±0,02	0,87±0,04	0,96±0,02	<b>0,97±0,02</b>	<b>0,99±0,01</b>

Таблица 7. Результаты тестирования обученных нейронных сетей и алгоритмов — значения F<sub>1</sub>-score относительно класса «шум»

Нейронная сеть (алгоритм)	Размер образца				
	5	10	20	30	50
MLP	0,84±0,01	0,38±0,46	0,92±0,01	0,91±0,01	0,94±0,02
FCN	0,84±0,02	<b>0,90±0,02</b>	<b>0,96±0,01</b>	0,96±0,01	0,98±0,01
ResNet	<b>0,85±0,02</b>	0,88±0,01	<b>0,96±0,01</b>	<b>0,97±0,01</b>	<b>0,98±0,00</b>
InceptionTime	0,84±0,02	0,90±0,03	<b>0,96±0,01</b>	0,96±0,01	0,98±0,01
TSTransformer	0,83±0,01	0,69±0,4	0,87±0,02	0,87±0,01	0,87±0,03
Random Forest	0,82±0,02	0,88±0,01	0,91±0,02	0,90±0,01	0,92±0,02
HIVE-COTE v2	0,68±0,02	0,82±0,01	0,92±0,02	0,93±0,02	0,98±0,01

Таблица 8. Результаты тестирования обученных нейронных сетей и алгоритмов — значения F<sub>1</sub>-score относительно класса «с возмущением»

Нейронная сеть (алгоритм)	Размер образца				
	5	10	20	30	50
MLP	0,81±0,02	0,74±0,13	0,91±0,01	0,91±0,01	0,94±0,02
FCN	0,84±0,02	<b>0,90±0,03</b>	<b>0,96±0,01</b>	0,96±0,01	0,98±0,01
ResNet	<b>0,85±0,03</b>	0,87±0,03	<b>0,96±0,01</b>	<b>0,97±0,01</b>	<b>0,98±0,00</b>
InceptionTime	0,84±0,02	0,89±0,04	<b>0,96±0,01</b>	0,96±0,01	0,98±0,01
TSTransformer	0,81±0,02	0,81±0,1	0,85±0,03	0,85±0,02	0,85±0,05
Random Forest	0,81±0,02	0,88±0,01	0,91±0,02	0,90±0,01	0,92±0,02
HIVE-COTE v2	0,70±0,02	0,84±0,01	0,92±0,02	0,94±0,02	0,98±0,01

Таблица 9. Результаты измерения времени вывода нейронных сетей на тестовых выборках

Нейронная сеть	Размер образца				
	5	10	20	30	50
	Время, мс (кол-во параметров)	Время, мс (кол-во параметров)	Время, мс (кол-во параметров)	Время, мс (кол-во параметров)	Время, мс (кол-во параметров)
MLP	<b>135,80±7,65</b> (8834)	<b>124,05±6,40</b> (9154)	<b>105,23±6,37</b> (9794)	<b>91,57±6,19</b> (10 434)	<b>70,84±10,50</b> (11 714)
FCN	209,58±8,97 (4065)	191,23±9,57 (8000)	154,36±14,05 (15 870)	136,67±6,34 (23 740)	108,75±12,20 (39 480)
ResNet	348,89±11,43 <b>(3684)</b>	312,21±12,59 <b>(3684)</b>	276,66±12,58 (12 122)	243,80±10,00 (12 122)	192,89±6,93 (12 122)
InceptionTime	502,92±17,02 (4265)	455,99±16,31 (4265)	387,06±12,44 <b>(4265)</b>	340,31±16,78 <b>(4265)</b>	265,47±9,54 <b>(4265)</b>
TSTransformer	372,06±13,12 (9842)	342,94±11,88 (10 002)	309,94±10,75 (10 322)	270,53±16,34 (10 642)	212,39±6,71 (11 282)

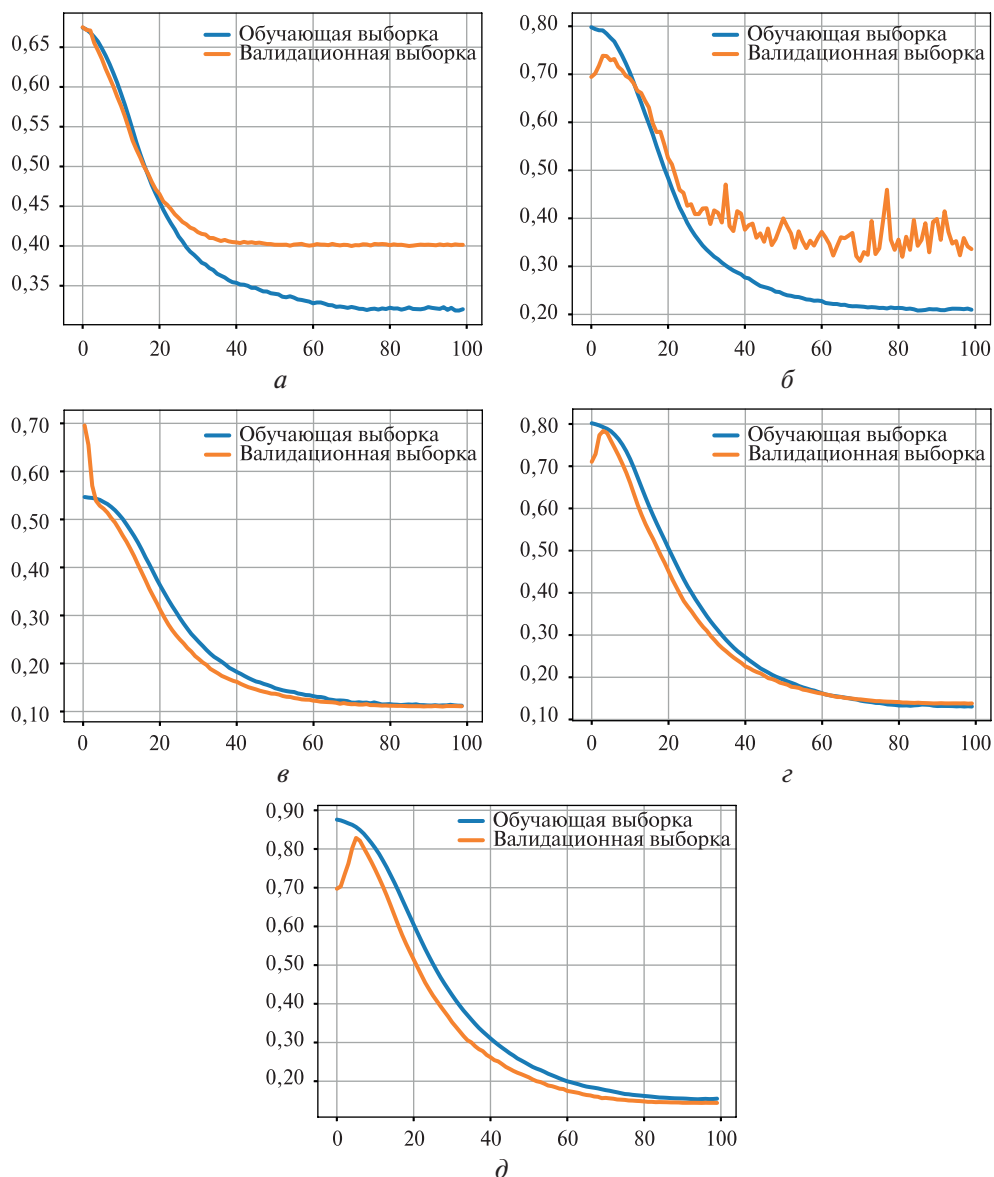


Рис. 7. Графики обучения нейронной сети InceptionTime, размер образца: а — 5; б — 10; в — 20; г — 30; д — 50; Для каждого графика: ось ординат — значение функции потерь, ось абсцисс — номер эпохи обучения

На рис. 7 (см. с. 47) представлены графики обучения для нейронной сети архитектуры InceptionTime. Для наборов данных с размером образцов 5 и 10 наблюдались признаки переобучения (расхождение кривых обучения), случившегося, вероятно, вследствие малого размера образцов. Графики обучения на наборах данных с образцами длиной 20, 30 и 50 показывают хорошую сходимость моделей.

Задачу поиска КИВ можно формулировать как задачу поиска аномалий: искомые возмущения встречаются редко среди основного массива шума и, вероятно, будет более естественным использовать для обучения нейронных сетей набор данных с таким же соотношением классов образцов. Из этих соображений был дополнительно сгенерирован несбалансированный набор данных с размером образца 30, количество образцов с возмущением — 1414, с шумом — 84 987, общее число образцов — 86 401. На этом наборе были обучены нейронные сети архитектур, показавших лучший результат на сбалансированных наборах данных: InceptionTime, ResNet, FCN. Обучение проводилось также с применением k-fold кросс-валидации. Метрика assuarcy непригодна для обучения на несбалансированных наборах данных в силу дисбаланса классов, поэтому использовались метрики  $F_1$ -мера, precision и recall. Значения метрик представлены в табл. 10 (приведены усреднённые значения со стандартным отклонением).

Таблица 10. Метрики качества классификации для несбалансированного набора данных

Нейронная сеть	Метрики (для класса с возмущением)		
	precision	recall	$F_1$ -score
ResNet	0,95±0,02	0,95±0,01	0,95±0,008
FCN	0,95±0,02	0,95±0,01	0,95±0,01
InceptionTime	0,94±0,01	0,95±0,01	0,94±0,006

Предварительные тесты предложенного алгоритма детектирования КИВ с размером окна 30 проведены для 1000 тестовых файлов с размеченными КИВ из архива за 15 июня 2009 г. Результаты экспериментов показаны в табл. 11. Там же представлены результаты работы алгоритма STA/LTA, который обрабатывает файл целиком, поэтому указанное количество ложных срабатываний понимается как на общее число файлов. Предложенный же алгоритм с классификатором нейронной сетью сначала извлекает из файлов образцы, а затем их классифицирует, поэтому в этом случае количество ложных срабатываний считается на общее число извлечённых образцов (84 567).

Таблица 11. Сравнение алгоритма STA/LTA и алгоритма на основе классификатора нейронной сети (разные архитектуры) — результаты сканирования тестовых размеченных данных

Алгоритм	Количество найденных КИВ (из 49)	Количество ложных срабатываний	Время работы, мс
STA/LTA	26	1163	1 900
Предложенный алгоритм (InceptionTime)	47	9482	12 228
Предложенный алгоритм (ResNet)	41	3475	9 494
Предложенный алгоритм (FCN)	47	7766	6 748
Предложенный алгоритм (InceptionTime, несбалансированный набор данных)	0	0	—
Предложенный алгоритм (ResNet, несбалансированный набор данных)	14	90	10 206
Предложенный алгоритм (FCN, несбалансированный набор данных)	5	1	7 113

При измерении времени работы алгоритмов не учитывалось время на подготовку данных (чтение файлов, извлечение образцов), замерялось только время выполнения основной части — вычисление значений STA/LTA либо вывод классификатора нейронной сети.

Предложенный алгоритм показывает высокую долю ложных срабатываний. Например, при сканировании тестовых файлов алгоритмом с нейронной сетью архитектуры InceptionTime, обученной на сбалансированном наборе данных (из файлов за предыдущие сутки), количество ложных срабатываний составляет 9482 (на общее количество в 84 567 образцов, извлечённых из сканируемых файлов), что существенно больше чем у алгоритма STA/LTA, но при этом найдено 47 из 49 реальных размеченных КИВ. Схожий результат показывает алгоритм с использованием нейронной сети архитектуры FCN.

Существенно меньше ложных срабатываний даёт алгоритм с классификатором нейронной сетью ResNet, обученной на несбалансированном наборе данных: всего 90 на 84 567 извлечённых образцов, но обнаружено всего 14 из 49 размеченных КИВ.

Значимой и непреодоленной в рамках данной работы проблемой оказалась неустойчивость обучения нейронных сетей на несбалансированных наборах данных. После обучения с k-fold кросс-валидацией получаемые пять экземпляров обученных нейронных сетей, несмотря на высокие значения метрик в тестах алгоритма, выдавали разный результат: например, ни один экземпляр нейронной сети InceptionTime, применённый в алгоритме, не находил ни одного КИВ. При этом на графиках обучения не наблюдалось признаков переобучения. Другие сети, согласно *табл. 12*, дают лучший результат.

Для создания универсального алгоритма необходимо, чтобы нейронная сеть, используемая как классификатор, обладала большой обобщающей способностью. Она должна хорошо отсеивать шум и разного рода иные возмущения, вызванные солнечной активностью, геомагнитными бурями и прочими факторами, и при этом с высокой точностью обнаруживать искомые ковулканические возмущения. Не менее важна и возможность применения алгоритма в исследованиях других событий, разделённых по времени и географическому положению. Поэтому в дальнейшем желательно использовать для обучения данные ПЭС других ГНСС-сетей и эруптивных событий, расположенных в разных точках мира.

Перспективным направлением исследования представляется дальнейшее использование несбалансированных наборов данных для обучения нейронных сетей вместе с применением техники весов классов, а также включение дополнительных условий поиска непосредственно в алгоритм, например величины минимальной и максимальной амплитуды и частоты полезного сигнала, продолжительности временного интервала, на котором осуществляется поиск КИВ. Рассмотренные нейронные сети при классификации не учитывают контекст текущего образца выборки, например тот факт, что амплитуда КИВ, как правило, превышает среднюю амплитуду фонового шума. Включение в работу алгоритма этого фактора должно помочь уменьшить количество ложных срабатываний.

## Заключение

В работе рассмотрена актуальная задача поиска ковулканических возмущений, ассоциированных с извержением влк. Пик Сарычева 14–15 июня 2009 г., на примере данных ГНСС-наблюдений. Показано, что предложенные архитектуры нейронных сетей в целом применимы для автоматизированного поиска КИВ. Наилучший результат для сбалансированного набора данных с размером образца 30 показали нейронные сети архитектур InceptionTime и FCN, обеспечивающие значение метрики assiguasy выше 0,95 и обнаружение в 1000 тестовых сериях ПЭС 47 из 49 КИВ. С точки зрения ресурсозатратности работы оптимальной выступает сеть архитектуры InceptionTime с 4265 параметрами. Необходимо отметить, что все рассмотренные нейронные сети, обученные на сбалансированных наборах данных, показывают большое количество ложных срабатываний алгоритма, а обучение нейронных сетей на несбалансированных данных, хоть и даёт результат с меньшим количеством ложных срабатываний, но на данном этапе нестабильно. Эти сложности требуют доработки предложенного алгоритма обработки данных. Тем не менее, по нашему мнению, полученные результаты убедительно

показывают, что при достижении достаточной обобщающей способности и высокого качества классификации нейронные сети пригодны для поиска ковулканических возмущений в рядах ПЭС как в апостериорном режиме, так и в режиме реального времени.

Авторы выражают благодарность консорциуму UNAVCO за предоставленные данные сети KurilNet. Исследования проведены с использованием ресурсов Центра коллективного пользования научным оборудованием «Центр обработки и хранения научных данных ДВО РАН» (Sorokin et al., 2017), финансируемого Министерством науки и высшего образования Российской Федерации (Минобрнауки) по соглашению № 075-15-2021-663. Работа выполнена при частичной поддержке гранта Дальневосточного федерального университета № 22-07-01-007 и государственного задания Минобрнауки России № 075-00771-22-00 (научно-исследовательские и опытно-конструкторские работы Института прикладной математики ДВО РАН № АААА-А20-120120390006-0).

## Литература

1. *Ahsan M. M., Mahmud M. A. P., Saha P. K., Gupta K. D., Siddique Z.* Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance // *Technologies*. 2021. V. 9. No. 3. P. 52. DOI: 10.3390/technologies9030052.
2. *Aji B. A. S., Liang T. H., Muslim B.* Detection precursor of Sumatra earthquake based on ionospheric total electron content anomalies using N-Model Artificial Neural Network // *Intern. Conf. Advanced Computer Science and Information Systems (ICACSIS)*. 2017. P. 269–276. DOI: 10.1109/ICACSIS.2017.8355045.
3. *Astafyeva E., Zakharenkova I. E., Förster M.* Ionospheric response to the 2015 St. Patrick's Day storm: A global multi-instrumental overview // *J. Geophysical Research: Space Physics*. 2015. V. 120. P. 9023–9037. DOI: 10.1002/2015JA021629.
4. *Azar J., Makhoul A., Couturier R., Demerjian J.* Robust IoT time series classification with data compression and deep learning // *Neurocomputing*. 2020. V. 398. P. 222–234. DOI: 10.1016/j.neucom.2020.02.097.
5. *Breiman L.* Random forests // *Machine Learning*. 2001. V. 45. No. 1. P. 5–32. DOI: 10.1023/A:1010950718922.
6. *Brissaud Q., Astafyeva E.* Near-real-time detection of co-seismic ionospheric disturbances using machine learning // *Geophysical J. Intern.* 2022. V. 230. No. 3. P. 2117–2130. DOI: 10.1093/gji/ggac167.
7. *Faouzi J., Janati H.* pyts: A python package for time series classification // *J. Machine Learning Research*. 2020. V. 21. No. 46. P. 1–6.
8. *Galvan D. A., Komjathy A., Hickey M. P., Stephens P., Snively J., Tony Song Y., Butala M. D., Mannucci A. J.* Ionospheric signatures of Tohoku-Oki tsunami of March 11, 2011: Model comparisons near the epicenter // *Radio Science*. 2011. V. 47. No. 4. DOI: 10.1029/2012RS005023.
9. *Han L., Wong J., Bancroft J.* Time picking on noisy microseismograms // *Proc. GeoCanada 2010 Convention — Working with the Earth*. 2010. 4 p.
10. *Heki K.* Explosion energy of the 2004 eruption of the Asama Volcano, central Japan, inferred from ionospheric disturbances // *Geophysical Research Letters*. 2006. V. 33. No. 14. DOI: 10.1029/2006GL026249.
11. *Fawaz I. H., Forestier G., Weber J., Idoumghar L., Muller P. A.* Deep learning for time series classification: a review // *Data Mining and Knowledge Discovery*. 2019. V. 33. No. 4. P. 917–963. DOI: 10.1007/978-3-030-88113-9\_24.
12. *Fawaz I. H., Lucas B., Forestier G., Pelletier C., Schmidt D. F., Weber J., Petitjean F.* InceptionTime: Finding AlexNet for time series classification // *Data Mining and Knowledge Discovery*. 2020. V. 34. No. 6. P. 1936–1962. DOI: 10.1007/s10618-020-00710-y.
13. *Kogan M. G., Vasilenko N. F., Frolov D. I., Freymueller J. T., Steblov G. M., Prytkov A. S., Ekström G.* Rapid postseismic relaxation after the great 2006–2007 Kuril earthquakes from GPS observations in 2007–2011 // *J. Geophysical Research: Solid Earth*. 2013. V. 118. No. 7. P. 3691–3706. DOI: 10.1002/jgrb.50245.
14. *Krischer L., Megies T., Barsch R., Beyreuther M., Lecocq T., Caudron C., Wassermann J.* ObsPy: a bridge for seismology into the scientific Python ecosystem // *Computational Science and Discovery*. 2015. V. 8. No. 1. DOI: 10.1088/1749-4699/8/1/014003.
15. *Levin V. E., Bakhtiarov V. F., Titkov N. N., Serovetnikov S. S., Magus'kin M. A., Lander A. V.* Contemporary crustal movements (CCMs) in Kamchatka // *Izvestiya Physics of the Solid Earth*. 2014. V. 50. No. 6. P. 732–751. <https://doi.org/10.1134/S1069351314060044>.
16. *Li H., Liu J., Yang Z., Liu R. W., Wu K., Wan Y.* Adaptively constrained dynamic time warping for time series classification and clustering // *Information Sciences*. 2020. V. 534. P. 97–116. DOI: 10.1016/j.ins.2020.04.009.

17. *Lin J., Chiou J.* Detecting Total Electron Content Precursors Before Earthquakes by Examining Total Electron Content Images Based on Butterworth Filter in Convolutional Neural Networks // *IEEE Access*. 2020. V. 8. P. 110478–110494. DOI: 10.1109/ACCESS.2020.3001337.
18. *Löning M., Bagnall A., Ganesh S., Kazakov V., Lines J., Király F.* sktime: A Unified Interface for Machine Learning with Time Series // *ArXiv:1909.07872*. 2019. DOI: 10.48550/ARXIV.1909.07872.
19. *Middlehurst M., Large J., Flynn M., Lines J., Bostrom A., Bagnall A.* HIVE-COTE 2.0: A new meta ensemble for time series classification // *Machine Learning*. 2021. V. 110. No. 11–12. P. 3211–3243. DOI: 10.1007/s10994-021-06057-9.
20. *Muafiry I. N., Meilano I., Heki K., Wijaya D. D., Nugraha K. A.* Ionospheric Disturbances after the 2022 Hunga Tonga-Hunga Ha'apai Eruption above Indonesia from GNSS-TEC Observations // *Atmosphere*. 2022. V. 13. No. 10. Art. No. 1615. DOI: 10.3390/atmos13101615.
21. *Nguyen L. H., Joshi D. R., Clay D. E., Henebry G. M.* Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier // *Remote Sensing of Environment*. 2020. V. 238. Art. No. 111017. DOI: 10.1016/j.rse.2018.12.016.
22. *Shestakov N., Orlyakovskiy A., Perevalova N., Titkov N., Chebrov D., Ohzono M., Takahashi H.* Investigation of Ionospheric Response to June 2009 Sarychev Peak Volcano Eruption // *Remote Sensing*. 2021. V. 13. No. 4. P. 1–18. DOI: 10.3390/rs13040638.
23. *Shults K., Astafyeva E., Adourian S.* Ionospheric detection and localization of volcano eruptions on the example of the April 2015 Calbuco events // *J. Geophysical Research: Space Physics*. 2016. V. 121. No. 10. P. 10303–10315. DOI: 10.1002/2016JA023382.
24. *Sorokin A. A., Makogonov S. V., Korolev S. P.* The Information Infrastructure for Collective Scientific Work in the Far East of Russia // *Scientific and Technical Information Processing*. 2017. V. 44. No. 4. P. 302–304. DOI: 10.3103/S0147688217040153.
25. *Tsuji H., Hatanaka Y., Hiyama Y., Yamaguchi K., Furuya T., Kawamoto S., Fukuzaki Y.* Twenty-Year Successful Operation of GEONET // *Bull. Geospatial Information Authority of Japan*. 2017. V. 65. P. 19–44.
26. *Usmankhujayev S., Ibrokhimov B., Baydadaev S., Kwon J.* Time series classification with inceptionFCN // *Sensors*. 2022. V. 22. No. 1. DOI: 10.3390/s22010157.
27. *Wang Z., Yan W., Oates T.* Time series classification from scratch with deep neural networks: A strong baseline // *Proc. Intern. Joint Conf. Neural Networks*. 2017. V. 2017. P. 1578–1585. DOI: 10.1109/IJCNN.2017.7966039.
28. *Withers M., Aster R. C., Young C. J., Beiriger J. I., Harris M., Moore S. G., Trujillo J. R.* A comparison of select trigger algorithms for automated global seismic phase and event detection // *Bull. Seismological Society of America*. 1998. V. 88(1). P. 95–106. DOI: 10.1785/bssa0880010095.
29. *Zerveas G., Jayaraman S., Patel D., Bhamidipaty A., Eickhoff C.* A transformer-based framework for multivariate time series representation learning // *Proc. ACM SIGKDD Intern. Conf. Knowledge Discovery and Data Mining*. 2021. P. 2114–2124. DOI: 10.1145/3447548.3467401.
30. *Zewdie G. K., Valladares C., Cohen M. B., Lary D. J., Ramani D., Tsidu G. M.* Data-driven forecasting of low-latitude ionospheric total electron content using the random forest and LSTM machine learning methods // *Space Weather*. 2021. V. 19(6). DOI: 10.1029/2020SW002639.

## Application of machine learning methods for detection of covolcanic ionospheric disturbances by GNSS observations data

A. S. Ten<sup>1</sup>, N. V. Shestakov<sup>2,6</sup>, A. A. Sorokin<sup>1</sup>,  
N. N. Titkov<sup>3</sup>, M. Ohzono<sup>4</sup>, H. Takahashi<sup>5</sup>

<sup>1</sup> Computing Center FEB RAS, Khabarovsk 680000, Russia  
E-mail: alexander.s.ten@yandex.ru

<sup>2</sup> Institute of Applied Mathematics FEB RAS, Vladivostok 690041, Russia

<sup>3</sup> Kamchatka Branch of the Geophysical Survey RAS  
Petropavlovsk-Kamchatsky 683023, Russia  
E-mail: nik@emsd.ru

<sup>4</sup> Earthquake Research Institute, University of Tokyo, Tokyo, Japan  
E-mail: m.ohzono@sci.hokudai.ac.jp

<sup>5</sup> Institute of Seismology and Volcanology, Hokkaido University, Sapporo, Japan  
E-mail: hiroaki@sci.hokudai.ac.jp

<sup>6</sup> Far Eastern Federal University, Vladivostok 690092, Russia  
E-mail: shestakov.nv@dyfu.ru

The work is devoted to the evaluation of the possibility of using artificial neural networks to detect covolcanic ionospheric disturbances in the time series of the total electronic content obtained from GNSS observations. Using the example of the Sarychev Peak volcano eruption on June 11–16, 2009, instrumental GNSS data were processed and labeled, their datasets with different sample sizes and two classes — with and without disturbances — were generated. Based on this information, five artificial neural networks of various architectures used to solve the problem of time series classification were trained and investigated. The quality metrics of classification of neural networks are calculated and their comparison is carried out. The results of testing the proposed algorithm with different classifiers — neural networks that showed the best result, namely InceptionTime and ResNet, were obtained. The analysis and comparison of these results with each other and with the results of the STA/LTA algorithm by the number of found covolcanic disturbances, false positives and the speed of operation is carried out. The disadvantages of STA/LTA regarding the application to the data under study and the possibility of overcoming them in the proposed approach are described. Conclusions are drawn about the quality of classification of neural networks and their applicability as a classifier in the universal algorithm for the detection of covolcanic disturbances. The directions of future research on the topic under consideration are proposed.

**Keywords:** ionosphere, covolcanic disturbances, machine learning, artificial neural networks, remote sounding, GNSS

Accepted: 19.01.2022

DOI: 10.21046/2070-7401-2023-20-1-37-54

### References

1. Ahsan M. M., Mahmud M. A. P., Saha P. K., Gupta K. D., Siddique Z., Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance, *Technologies*, 2021, Vol. 9, No. 3, p. 52, DOI: 10.3390/technologies9030052.
2. Aji B. A. S., Liang T. H., Muslim B., Detection precursor of Sumatra earthquake based on ionospheric total electron content anomalies using N-Model Artificial Neural Network, *Intern. Conf. Advanced Computer Science and Information Systems (ICACSIS)*, 2017, pp. 269–276, DOI: 10.1109/ICACSIS.2017.8355045.
3. Astafyeva E., Zakharenkova I. E., Förster M., Ionospheric response to the 2015 St. Patrick's Day storm: A global multi-instrumental overview, *J. Geophysical Research: Space Physics*, 2015, Vol. 120, pp. 9023–9037, DOI: 10.1002/2015JA021629.
4. Azar J., Makhoul A., Couturier R., Demerjian J., Robust IoT time series classification with data compression and deep learning, *Neurocomputing*, 2020, Vol. 398, pp. 222–234, DOI: 10.1016/j.neucom.2020.02.097.
5. Breiman L., Random forests, *Machine Learning*, 2001, Vol. 45, No. 1, pp. 5–32, DOI: 10.1023/A:1010950718922.



6. Brissaud Q., Astafyeva E., Near-real-time detection of co-seismic ionospheric disturbances using machine learning, *Geophysical J. Intern.*, 2022, Vol. 230, No. 3, pp. 2117–2130, DOI: 10.1093/gji/ggac167.
7. Faouzi J., Janati H., pyts: A python package for time series classification, *J. Machine Learning Research*, 2020, Vol. 21, No. 46, pp. 1–6.
8. Galvan D.A., Komjathy A., Hickey M.P., Stephens P., Snively J., Tony Song Y., Butala M.D., Mannucci A.J., Ionospheric signatures of Tohoku-Oki tsunami of March 11, 2011: Model comparisons near the epicenter, *Radio Science*, 2011, Vol. 47, No. 4, DOI: 10.1029/2012RS005023.
9. Han L., Wong J., Bancroft J., Time picking on noisy microseismograms, *Proc. GeoCanada 2010 Convention — Working with the Earth*, 2010, 4 p.
10. Heki K., Explosion energy of the 2004 eruption of the Asama Volcano, central Japan, inferred from ionospheric disturbances, *Geophysical Research Letters*, 2006, Vol. 33, No. 14, DOI: 10.1029/2006GL026249.
11. Fawaz I.H., Forestier G., Weber J., Idoumghar L., Muller P.A., Deep learning for time series classification: a review, *Data Mining and Knowledge Discovery*, 2019, Vol. 33, No. 4, pp. 917–963, DOI: 10.1007/978-3-030-88113-9\_24.
12. Fawaz I.H., Lucas B., Forestier G., Pelletier C., Schmidt D.F., Weber J., Petitjean F., InceptionTime: Finding AlexNet for time series classification, *Data Mining and Knowledge Discovery*, 2020, Vol. 34, No. 6, pp. 1936–1962, DOI: 10.1007/s10618-020-00710-y.
13. Kogan M.G., Vasilenko N.F., Frolov D.I., Freymueller J.T., Steblov G.M., Prytkov A.S., Ekström G., Rapid postseismic relaxation after the great 2006–2007 Kuril earthquakes from GPS observations in 2007–2011, *J. Geophysical Research: Solid Earth*, 2013, Vol. 118, No. 7, pp. 3691–3706, DOI: 10.1002/jgrb.50245.
14. Krischer L., Megies T., Barsch R., Beyreuther M., Lecocq T., Caudron C., Wassermann J., ObsPy: a bridge for seismology into the scientific Python ecosystem, *Computational Science and Discovery*, 2015, Vol. 8, No. 1, DOI: 10.1088/1749-4699/8/1/014003.
15. Levin V.E., Bakhtiarov V.F., Titkov N.N., Serovetnikov S.S., Magus'kin M.A., Lander A.V., Contemporary crustal movements (CCMs) in Kamchatka, *Izvestiya Physics of Solid Earth*, 2014, Vol. 50, No. 6, pp. 732–751, <https://doi.org/10.1134/S1069351314060044>.
16. Li H., Liu J., Yang Z., Liu R.W., Wu K., Wan Y., Adaptively constrained dynamic time warping for time series classification and clustering, *Information Sciences*, 2020, Vol. 534, pp. 97–116, DOI: 10.1016/j.ins.2020.04.009.
17. Lin J., Chiou J., Detecting Total Electron Content Precursors Before Earthquakes by Examining Total Electron Content Images Based on Butterworth Filter in Convolutional Neural Networks, *IEEE Access*, 2020, Vol. 8, pp. 110478–110494, DOI: 10.1109/ACCESS.2020.3001337.
18. Löning M., Bagnall A., Ganesh S., Kazakov V., Lines J., Király F., sktime: A Unified Interface for Machine Learning with Time Series, *ArXiv:1909.07872*, 2019, DOI: 10.48550/ARXIV.1909.07872.
19. Middlehurst M., Large J., Flynn M., Lines J., Bostrom A., Bagnall A., HIVE-COTE 2.0: A new meta ensemble for time series classification, *Machine Learning*, 2021, Vol. 110, No. 11–12, pp. 3211–3243, DOI: 10.1007/s10994-021-06057-9.
20. Muafiry I.N., Meilano I., Heki K., Wijaya D.D., Nugraha K.A., Ionospheric Disturbances after the 2022 Hunga Tonga-Hunga Ha'apai Eruption above Indonesia from GNSS-TEC Observations, *Atmosphere*, 2022, Vol. 13, No. 10, Art. No. 1615, DOI: 10.3390/atmos13101615.
21. Nguyen L.H., Joshi D.R., Clay D.E., Henebry G.M., Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier, *Remote Sensing of Environment*, 2020, Vol. 238, Art. No. 111017, DOI: 10.1016/j.rse.2018.12.016.
22. Shestakov N., Orlyakovskiy A., Perevalova N., Titkov N., Chebrov D., Ohzono M., Takahashi H., Investigation of Ionospheric Response to June 2009 Sarychev Peak Volcano Eruption, *Remote Sensing*, 2021, Vol. 13, No. 4, pp. 1–18, DOI: 10.3390/rs13040638.
23. Shults K., Astafyeva E., Adourian S., Ionospheric detection and localization of volcano eruptions on the example of the April 2015 Calbuco events, *J. Geophysical Research: Space Physics*, 2016, Vol. 121, No. 10, pp. 10303–10315, DOI: 10.1002/2016JA023382.
24. Sorokin A.A., Makogonov S.V., Korolev S.P., The Information Infrastructure for Collective Scientific Work in the Far East of Russia, *Scientific and Technical Information Processing*, 2017, Vol. 44, No. 4, pp. 302–304, DOI: 10.3103/S0147688217040153.
25. Tsuji H., Hatanaka Y., Hiyama Y., Yamaguchi K., Furuya T., Kawamoto S., Fukuzaki Y., Twenty-Year Successful Operation of GEONET, *Bull. Geospatial Information Authority of Japan*, 2017, Vol. 65, pp. 19–44.
26. Usmankhujayev S., Ibrokhimov B., Baydadaev S., Kwon J., Time Series Classification with InceptionFCN, *Sensors*, 2022, Vol. 22, No. 1, DOI: 10.3390/s22010157.
27. Wang Z., Yan W., Oates T., Time series classification from scratch with deep neural networks: A strong baseline, *Proc. Intern. Joint Conf. Neural Networks*, 2017, Vol. 2017, pp. 1578–1585, DOI: 10.1109/IJCNN.2017.7966039.

28. Withers M., Aster R. C., Young C. J., Beiriger J. I., Harris M., Moore S. G., Trujillo J. R., A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. Seismological Society of America*, 1998, Vol. 88(1), pp. 95–106, DOI: 10.1785/bssa0880010095.
29. Zerveas G., Jayaraman S., Patel D., Bhamidipaty A., Eickhoff C., A transformer-based framework for multivariate time series representation learning, *Proc. ACM SIGKDD Intern. Conf. Knowledge Discovery and Data Mining*, 2021, pp. 2114–2124, DOI: 10.1145/3447548.3467401.
30. Zewdie G. K., Valladares C., Cohen M. B., Lary D. J., Ramani D., Tsidu G. M., Data-driven forecasting of low-latitude ionospheric total electron content using the random forest and LSTM machine learning methods, *Space Weather*, 2021, Vol. 19(6), DOI: 10.1029/2020SW002639.