

# Применение признаков на основе дискретного вейвлет-преобразования для обучения моделей машинного обучения в задаче классификации ковулканических ионосферных возмущений

А.С. Тен<sup>1</sup>, А.А. Сорокин<sup>1</sup>, Н.В. Шестаков<sup>2</sup>

<sup>1</sup>ВЦ ДВО РАН, Хабаровск

<sup>2</sup>ДВФУ \ ИПМ ДВО РАН, Владивосток

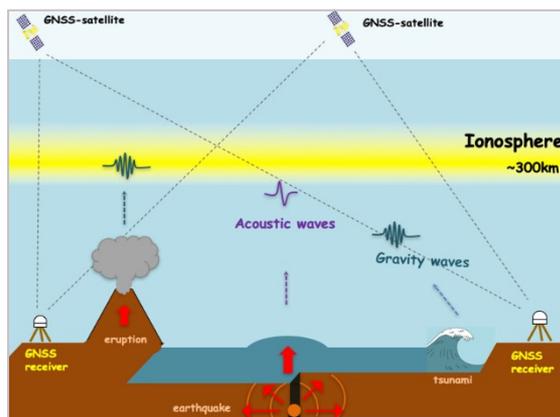
XXII международная конференция “СОВРЕМЕННЫЕ ПРОБЛЕМЫ  
ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА”

11 - 15 ноября 2024, г. Москва

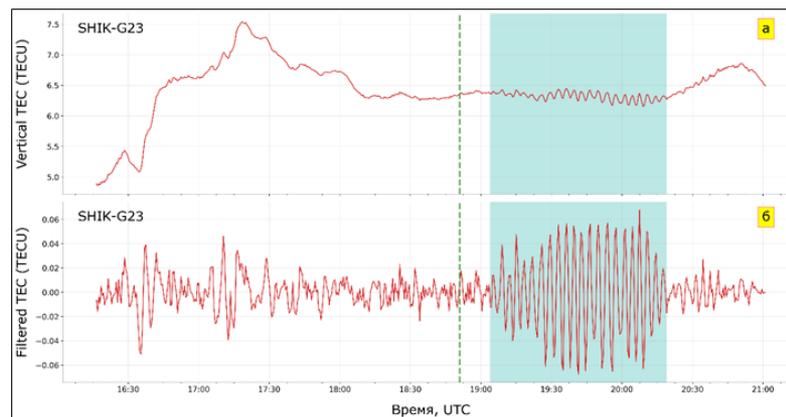


# Актуальность

- Извержения вулканов (VEI 2-6) порождают в ионосфере Земли ковулканические ионосферные возмущения (КИВ)<sup>1</sup>.
- Для исследования КИВ анализируются временные ряды полного электронного содержания (ПЭС).
- Ряды ПЭС реконструируют из данных глобальных навигационных спутниковых систем (ГНСС) – GPS, ГЛОНАСС, «Бэйдоу», «Галилео».



источник изображения: <https://eos.org/editors-vox/detecting-earths-natural-hazards-high-up-in-the-sky>



# Актуальность

- Большой объем данных ГНСС (тысячи файлов).
- Существующие методы поиска КИВ (STA\LTA, DROT<sup>1</sup>) не способны эффективно различать ковулканические возмущения и возмущения иной природы (рис. 1).
- Много вспомогательных операций с данными – высокий риск ошибок.



Рис. 1 Временной ряд ПЭС с КИВ и возмущением иной природы

# Возможные решения

- Классификация временных рядов ПЭС.
- Методы машинного обучения и нейронные сети для анализа временных рядов уже успешно применяются в сейсмологии<sup>1</sup>, метеорологии<sup>2</sup> и др. областях.
- Новые вычислители и адаптированный стек ПО.

1.Stepnov A., Chernykh V., Konovalov A. The seismo-performer: a novel machine learning approach for general and efficient seismic phase recognition from local earthquakes in real time // Sensors. 2021. № 18 (21).

2.Андреев А.И., Шамилова Ю.А. Детектирование облачности по данным КА Himawari-8 с применением сверточной нейронной сети // Исследование Земли из космоса. 2021. №2. с. 42-52.

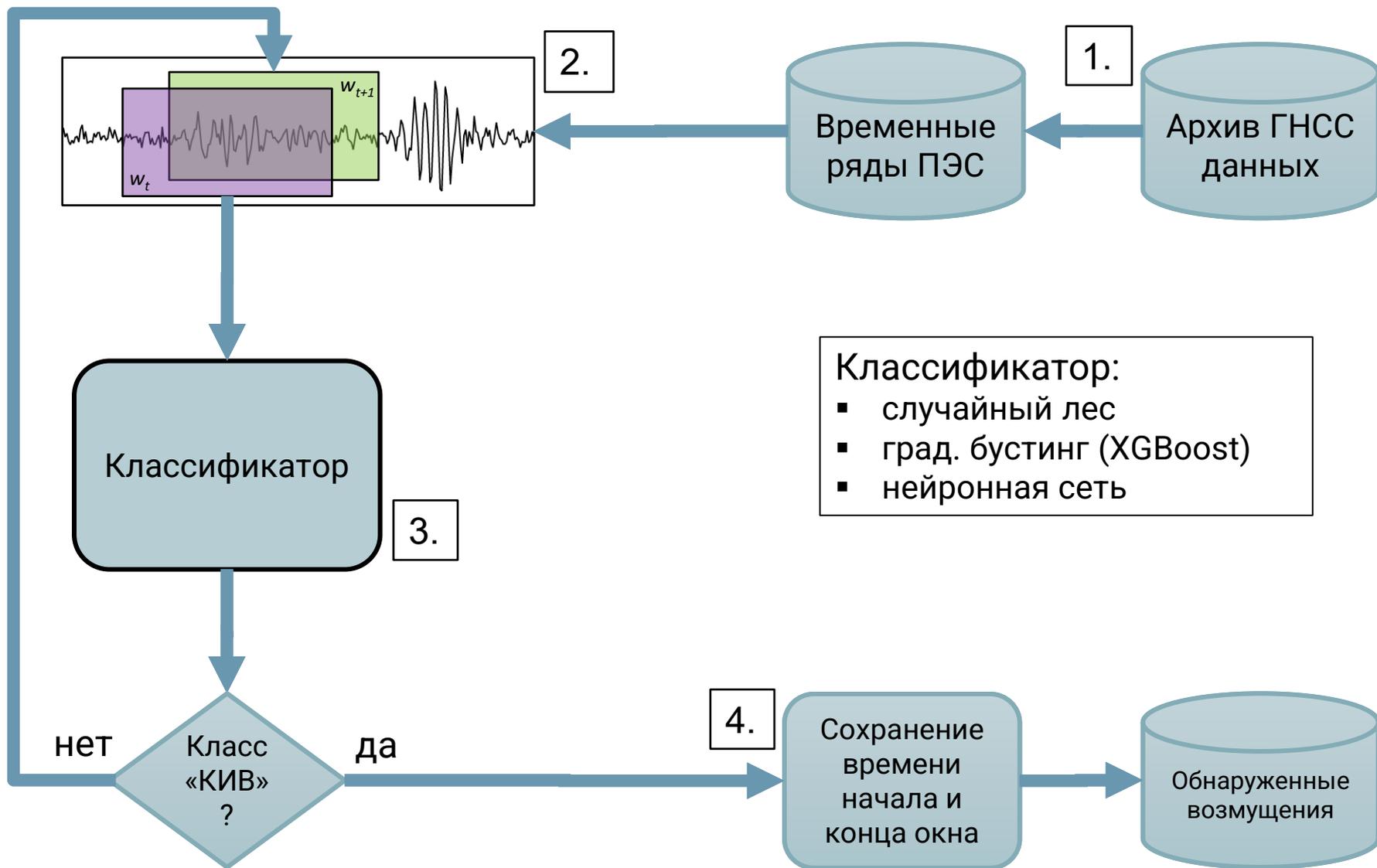
# Цель и задачи исследования

Оценка применимости и сравнение некоторых методов классического машинного обучения в задаче бинарной классификации волновых форм ПЭС.

Основные задачи:

1. Сформировать обучающие наборы данных.
2. Сконструировать эффективные признаки данных на основе дискретного вейвлет-преобразования исходных временных рядов.
3. Обучить выбранные модели машинного обучения, сравнить метрики качества классификации.
4. Сравнить обученные модели в алгоритме автоматизированного поиска КИВ на тестовых ГНСС-данных.

# Алгоритм поиска ковулканических ионосферных возмущений



# Классификаторы

Классификатор	Тип	Тип обучения	Данные для обучения
Случайный лес	Классическая модель машинного обучения	Обучение с учителем	Признаки
Градиентный бустинг (XGBoost)			
FCN <sup>1</sup>	Нейронная сеть	Обучение с учителем	Исходные временные ряды ПЭС
InceptionTime <sup>1</sup>			
PCA	Классическая модель машинного обучения	Обучение без учителя	Признаки
KNN			
Isolation Forest			

1. Тен А.С. [и др.]. Применение методов машинного обучения для поиска ковулканических ионосферных возмущений по данным ГНСС-наблюдений // Современные проблемы дистанционного зондирования Земли из космоса. — 2023. — Т. 20. — № 1. — С. 37-54.
2. Chen T., Guestrin C. XGBoost: A scalable tree boosting system KDD '16 / ACM, 2016
3. Fawaz H. I. [и др.]. Deep learning for time series classification: a review // Data Mining and Knowledge Discovery. 2018. (33). С. 917–963.
4. Ismail Fawaz H. [и др.]. InceptionTime: Finding AlexNet for time series classification // Data Mining and Knowledge Discovery. 2020. № 6 (34). С. 1936–1962.
5. Liu F. T., Ting K. M., Zhou Z.-H. Isolation-based anomaly detection // ACM Trans. Knowl. Discov. Data. 2012. № 1 (6).

# 1. Формирование обучающих наборов

Вулкан	Географическое расположение	Извержение	Индекс эксплозивности (VEI)	Количество станций	Источник данных (сеть ГНСС)	Количество реконструированных рядов ПЭС
Пик Сарычева <sup>1</sup>	Курилы, Россия	14-15 июня 2009 года, 18 ч 51 мин UT/ 01 ч 15 мин UTC	4	318	GEONET, IGS ( <a href="http://www.igs.org/">http://www.igs.org/</a> ), KurilNet, KamNet	18958
Кальбуко <sup>3</sup>	Анды, Чили	22 апреля 2015 г. в 21 ч. 04 мин UTC	4	25	IGS ( <a href="http://www.igs.org/">http://www.igs.org/</a> )	962
Хунга-Тонга-Хунга-Хаапай <sup>2</sup>	Архипелаг Тонга	15 января 2022 г. примерно в 4 ч. 02 мин по UTC	5	11	IGS ( <a href="http://www.igs.org/">http://www.igs.org/</a> )	336

1. Shestakov N. [и др.]. Investigation of ionospheric response to June 2009 Sarychev Peak volcano eruption // Remote Sensing. 2021. № 4 (13).

2. Abbrescia M. [и др.]. Observation of Rayleigh-Lamb waves generated by the 2022 Hunga-Tonga volcanic eruption with the POLA detectors at Ny-Ålesund // Scientific Reports. 2022. № 1 (12). С. 19978.

3. Shults K., Astafyeva E., Adourian S. Ionospheric detection and localization of volcano eruptions on the example of the April 2015 Calbuco events // Journal of Geophysical Research: Space Physics. 2016. № 10 (121). С. 10,303-10,315.

# 1. Формирование обучающих наборов

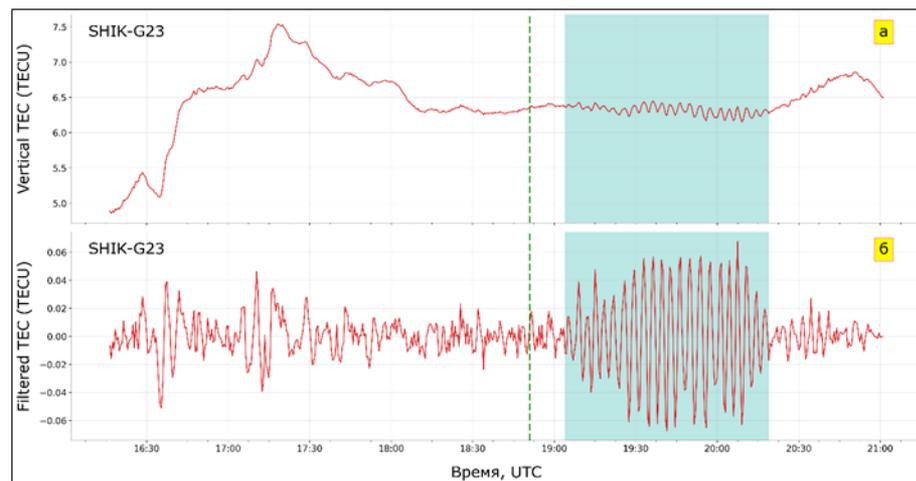
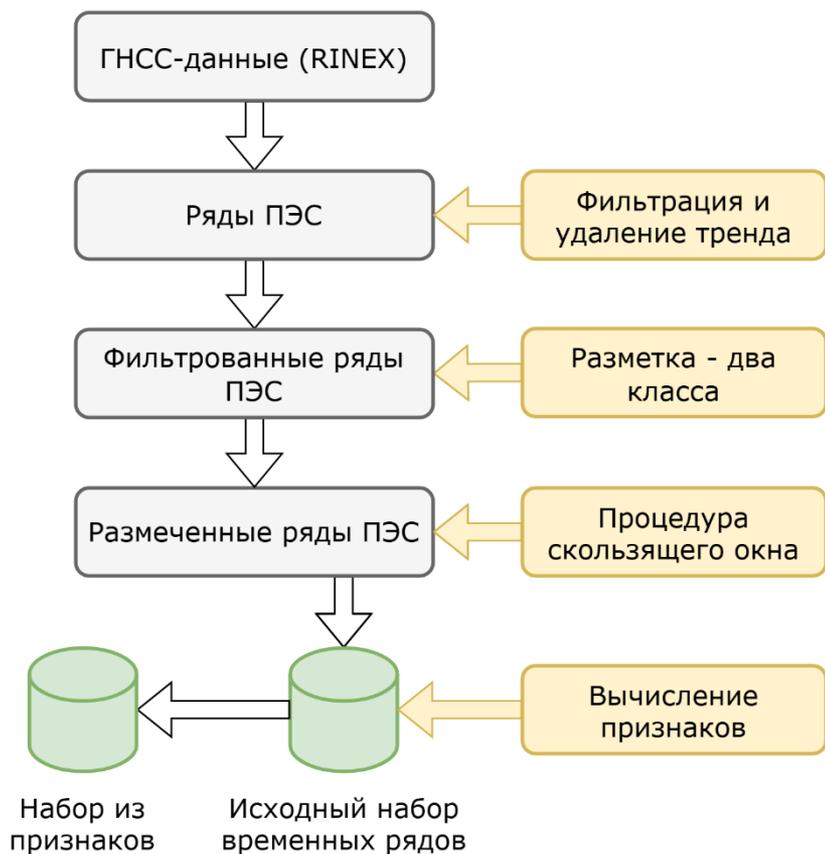


Рис. 2. Графики временных рядов ПЭС с размеченными ковулканическими возмущениями (вулкан Пик Сарычева, 14-15 июня 2009 г., а – график ряда вертикального ПЭС; б – график фильтрованных значений ПЭС.

Вертикальной пунктирной линией выделено время eruptивного события

Образцы 2х классов: «КИВ» и «шум».

## 2. Конструирование признаков временных рядов ПЭС на основе вейвлет-анализа

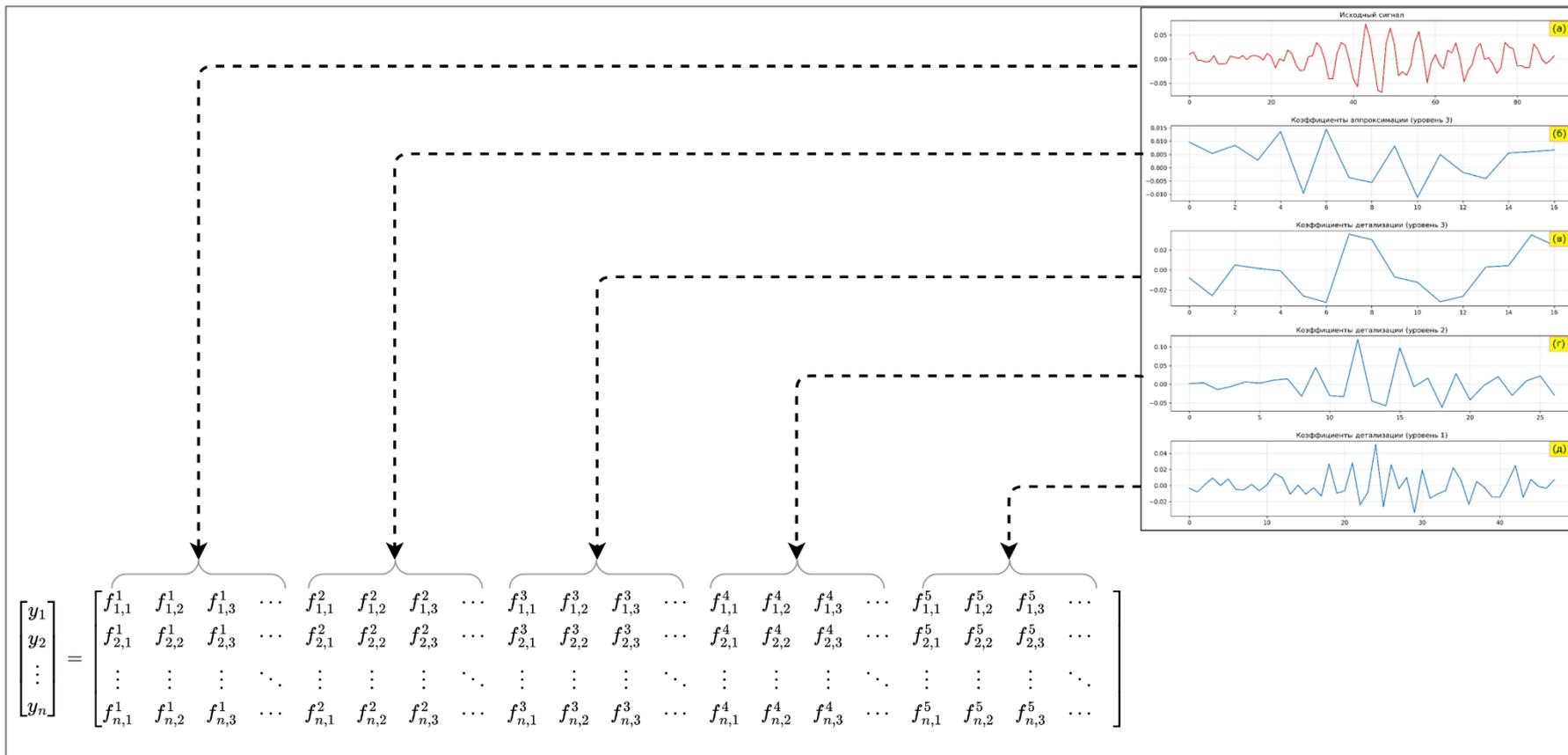
- Дискретное вейвлет-преобразование (DWT):
- является эффективным инструментом для построения признаков из временных рядов [1, 2].
- позволяет разложить временной ряд на набор коэффициентов, которые характеризуют его структуру на различных временных и частотных масштабах.
- позволяет выделить высокочастотные компоненты, отражающие краткосрочные изменения, так и низкочастотные, отвечающие за долгосрочные тенденции.

Использовался вейвлет “db4” – из семейства Добеши (Daubechies).

1. Li D. [и др.]. Time Series Classification with Discrete Wavelet Transformed Data // International Journal of Software Engineering and Knowledge Engineering. 2016. № 09n10 (26). С. 1361–1377.

2. Yan L., Liu Y., Liu Y. Application of Discrete Wavelet Transform in Shapelet-Based Classification // Mathematical Problems in Engineering. 2020. № 1 (2020). С. 6523872.

## 2. Конструирование признаков временных рядов ПЭС на основе вейвлет-анализа



$f_{n,j}^i$  – значение  $j$ -го признака на  $i$ -м уровне,  $n$ -го семпла,

$y_n$  – метка  $n$ -го семпла (0 = шум, 1 = КИВ),

$n$  – кол-во семплов в выборке.

Признаки ( $j$ ):

- среднее значение, медиана,
- стандартное отклонение, дисперсия,
- среднеквадратическое значение,
- коэффициент асимметрии, коэффициент эксцесса,
- спектральная энтропия,
- $k$ -е перцентили (5-й, 25-й, 75-й и 95-й).

### 3. Обучение классификаторов и оценка качества классификации

Модель \ метрика	Обучающая часть		Тестовая часть	
	F1	MCC	F1	MCC
Случайный лес	0.88±0.02	0.76±0.04	0.91	0.81
XGBoost	0.89±0.01	0.78±0.02	0.88	0.77
FCN	0.93±0.01	0.91±0.01	0.85	0.72
InceptionTime	0.93±0.01	0.91±0.01	0.85	0.74
PCA	-	-	0.61	0.34
KNN	-	-	0.58	0.25
Isolation Forest	-	-	0.61	0.33

### 3. Обучение классификаторов и оценка качества классификации

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$F_1 = 2 \frac{recall \cdot precision}{recall + precision} = \frac{2TP}{TP + (FP + FN)}$$

$$precision = \frac{TP}{TP + FP}, \quad recall = \frac{TP}{TP + FN}$$

		Predicted	
		Negative	Positive
Actual	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

## 4. Оценка эффективности классификаторов в алгоритме поиска КИВ

Событие	Всего файлов	КИВ содержится	Алгоритм/классификатор	КИВ, найдено	Ложных срабатываний, всего (в среднем на файл)
Хунга Тонга-Хунга Хаапай, 2022	336	32	Случайный лес	6 (18.75%)	28 (0,083)
			XGBoost (вейвлет)	22 (68.75%)	49 (0,15)
			XGBoost (ДПФ) <sup>1</sup>	26 (81.25%)	90 (0.27)
			FCN	27 (84%)	58 (0,17)
			InceptionTime	23 (71,9%)	39 (0,12)
			PCA	32 (100%)	295 (0,88)
			KNN	32 (100%)	299 (0,89)
			Isolation Forest	29 (90.63%)	296 (0,88)
Кальбуко, 2015	962	100	Случайный лес	35 (35%)	45 (0,047)
			XGBoost (вейвлет)	61 (61%)	63 (0,065)
			XGBoost (ДПФ) <sup>1</sup>	94 (94%)	164 (0.17)
			FCN	98 (98%)	213 (0,22)
			InceptionTime	97 (97%)	178 (0,12)
			PCA	99 (99%)	641 (0,67)
			KNN	99 (99%)	720 (0,75)
			Isolation Forest	99 (99%)	586 (0,61)
			STA\LTА	53 (53%)	866 (0,9)

1. Ten A., Sorokin A., Shestakov N., Ohzono M., Titkov N. Detecting covolcanic ionospheric disturbances using GNSS data and a machine learning algorithm // Advances in Space Research. – 2024.

2. x64-based PC, AMD Ryzen 7 4700U 8 cores 2.00 GHz, 16Gb RAM, 5 workers

# Заключение

- Признаки, основанные на дискретном вейвлет преобразовании, позволяют достигать приемлемого качества бинарной классификации временных рядов ПЭС и применять обученные на них модели в предложенном алгоритме автоматизированного поиска КИВ.
- Лучший результат – классификатор на основе градиентного бустинга, по соотношению числа найденных КИВ и ложных срабатываний на тестовых данных.
- Модели обучения без учителя могут применяться для задачи предразметки данных.
- Улучшение качества классификации и точности алгоритма поиска КИВ – отбор признаков, увеличение выборки данных с КИВ, использование дополнительной информации о солнечной активности, магнитных бурях и пр.

# Спасибо за внимание!

Александр Тен,  
м.н.с. лаборатории информационных  
и вычислительных систем ВЦ ДВО РАН  
E-mail: [alexander.s.ten@yandex.ru](mailto:alexander.s.ten@yandex.ru)

