

ПРОГНОЗИРОВАНИЕ СОЛНЕЧНЫХ ВСПЫШЕК И ФОНОВЫХ ПОТОКОВ РЕНТГЕНОВСКОГО ИЗЛУЧЕНИЯ ПО ДАННЫМ СИНОПТИЧЕСКИХ НАЗЕМНЫХ НАБЛЮДЕНИЙ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

© 2020 г. А. Г. Тлатов^{1, *}, Е. А. Илларионов², И. А. Березин¹, А. Д. Шрамко¹

¹Кисловодская горная астрономическая станция РАН, г. Кисловодск, Россия

²Московский государственный университет им. М.В. Ломоносова, г. Москва, Россия

*tlatov@mail.ru

Поступила в редакцию 01.03.2020 г.

После доработки 24.04.2020 г.

Принята к публикации 29.05.2020 г.

В работе представлены модели машинного обучения для прогнозирования мощных солнечных вспышек и фоновых потоков рентгеновского излучения в диапазоне 1–8 Å. Для прогнозирования солнечных вспышек на следующий день использовалась информация о текущем уровне солнечной активности, получаемая с наземных синоптических наблюдений, таких как характеристики солнечных пятен, потоки радиоизлучения на длинах волн 10.7 и 5 см, а также уровень фонового потока и количества солнечных вспышек текущего дня, полученных со спутника *GOES*. Для прогнозирования фоновых потоков рентгеновского излучения использовались только данные наземных телескопов. Показана высокая эффективность прогноза на следующий день. Нейронная сеть обучалась на данных, доступных с 2002 г.

DOI: 10.31857/S0023420620060102

ВВЕДЕНИЕ

В настоящее время прогнозирование космической погоды (КП), вызываемой солнечной активностью, является необходимым условием для успешного осуществления космических программ, авиасообщения на высоких широтах и других задач, в том числе и задач специального назначения. Для эффективного прогнозирования параметров космической погоды необходимы также прогнозы солнечных вспышек и уровня рентгеновского излучения. В данной работе представлены модели машинного обучения, решающие эти проблемы.

В данной работе мы представляем модели для прогнозирования солнечных вспышек на основе ежедневных синоптических наблюдений солнечной активности в оптическом и радио-диапазонах. Первая многофакторная модель построена на основе машинного обучения. Модель позволяет осуществлять прогноз количества и мощности солнечных вспышек за один-два дня с высокой достоверностью. Вторая, модель для прогнозирования фонового потока рентгеновского излучения по данным наземных наблюдений на основе моделей машинного обучения.

Модели построены на данных, включающие данные “классических” оптических телескопов

для наблюдения фотосферы, хромосферы и короны Солнца.

МЕТОДЫ ПРОГНОЗИРОВАНИЯ И ИСХОДНЫЕ ДАННЫЕ

Механизм солнечных вспышек является одной из основных нерешенных задач в физике Солнца. Процессы накопления энергии и запуска вспышек обусловлены появлением потока магнитного поля в фотосфере (например [1, 2]). Форма и сложность солнечных пятен в излучении белого света были классифицированы в соответствии с уровнем роста солнечных пятен [3]. Опытным путем известно, что более крупные пятна с большим количеством ядер и более сложной структурой магнитного потока имеют тенденцию вызывать более крупные вспышки (см., например, [4–9], а также повторные вспышки в одних и тех же активных областях (AR)) [10].

Одним из направлений методов прогнозирования является изучение характеристик отдельных активных областей. Так широко используются характеристики, полученные из магнитограмм, такие как магнитный поток, градиент магнитного поля [11], длина магнитных нейтральных линий [12], “эффективное” магнитное поле [13], магнитный

поток вблизи нейтральных магнитных линий [14, 15], общее магнитное поле диссипация энергии [16], взвешенная длина магнитной нейтральной линии и расстояние между центрами солнечных пятен с полярностью NS [17], непотенциальность [18] и вейвлет-спектры [19]. Эти особенности связаны с динамикой возникновения нового магнитного потока и тесно связаны с накоплением энергии и механизмами запуска.

Количество данных наблюдений за Солнцем в открытом доступе настолько велико, что не поддается обработке человеком. Чтобы справиться с проблемой прогнозирования вспышек, были применены различные алгоритмы машинного обучения: нейронная сеть [20–22], деревья решений [23], квантование вектора обучения [23], регрессионная модель [16], векторная машина [24], регрессия [25] и ансамбль из четырех предикторов [26]. В этих моделях часто основными данными являются данные магнитографических наблюдений. Вместе с тем оперативные данные наблюдения магнитографов зарубежных телескопов у нас не доступны или представлены не калиброванными. Это значительно снижает возможности для оперативного прогнозирования. Необходимо создание системы прогнозирования, основанной на отечественных и/или доступных зарубежных данных наблюдений.

В данной работе мы разработали модель прогнозирования вспышек с использованием машинного обучения, которая оптимизирована для прогнозирования вспышек максимального балла класса M и X , происходящих на следующий день. Машинное обучение используется для разработки алгоритмов, которые могут учиться и принимать решения по большому количеству данных. Мы использовали данные синоптических солнечных наблюдений за период 2002–2019 гг. К их числу относятся: наблюдения солнечных пятен. Это индекс солнечных пятен R , общая площадь пятен S , а также площадь наибольшей активной области S_{\max}^{gr} и площадь максимального пятна $S_{\text{spot}}^{\text{max}}$, поток радиоизлучения на волнах 10.7 см (R_{10}) и 5 см (R_5). Поток фонового рентгеновского излучения в диапазоне 1–8 Å, индекс солнечных вспышек за текущий и предыдущий день Fi . Индекс солнечных вспышек Fi мы определяли согласно [27]. Часть данных находится в базах данных Кисловодской горной астрономической станции (<http://solarstation.ru/>). Индекс солнечных пятен R вычисляется в соответствии с числом солнечных пятен Вольфа $R = k(10g + s)$, где g – количество групп солнечных пятен, s – общее количество отдельных пятен во всех группах, а k – коэффициент, который рассчитывается для каждой обсерватории. Индекс площади солнечных пятен S , который является суммой скорректированной площади всех наблюдаемых солнечных пятен в

миллионных долях солнечного полушария. Индекс потока радиоизлучения R_5 на волне 5 см, который определяется ежедневно в процессе калибровки. Другими данными являются данные ежедневных данных солнечной активности из базы данных ftp://ftp.swpc.noaa.gov/pub/indices/old_indices/. Это данные фонового потока рентгеновских лучей Xbg : среднесуточный фоновый поток рентгеновского излучения, измеренный на основном спутнике *GOES SWPC*. Алгоритм вычисления ежедневного рентгеновского фонового потока Xbg следующий. Изменения в диапазоне 1–8 Å в течение суток разделяются на 3 8-часовых интервала. Для каждого из этих восьми 8-часовых участков определяется минимум: $\min1$, $\min2$ и $\min3$. Далее рассчитывается среднее значение минимумов из первого и последнего 8-часовых отрезков: $\min_{\text{avg}} = 1/2(\min1 + \min3)$. Дневной фоновый поток рентгеновских лучей Xbg является наименьшим из $\min2$ или \min_{avg} . Также по этой базе данных мы вычисляли индекс солнечных вспышек: $Fi = 0.1N_B + 1N_C + 10N_M + 100N_X$, где N_B , N_C , N_M , N_X – количество вспышек классов B , C , M , X в текущем дне соответственно.

Для машинного обучения мы использовали сверточную нейронную сеть и плотно-разреженно-плотную схему обучения (*Dense-Sparse-Dense Training, DSD*, [28]). Обучение по схеме *DSD* состоит из 3 последовательных шагов. На первом шаге (*Dense*) происходит обучение нейронной сети с учетом всех данных. На следующем шаге (*Sparse*) часть весов каждого слоя зануляется и происходит обучение по оставшимся весам. Для зануления выбирается определенный процент весов сети (обычно 25–50%) ближайших к 0 по абсолютной величине. Выбранные веса продолжают оставаться нулевыми на каждой итерации обучения шага *Sparse*. На последнем шаге (*Dense*) включаются ранее сокращенные веса из предыдущего шага, и сеть обучается снова. Как показано в оригинальной работе [28], подобная схема улучшает точность обучаемой модели по сравнению с нормальной схемой обучения. Для оценки значимости входных признаков на результат прогноза мы обучили вспомогательную модель, состоящую из ансамбля решающих деревьев [29].

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Для обучения мы использовали все данные ежедневных наблюдений из интервала 2002–2018 гг. за исключением 2017 г. Для оценки достоверности модели мы использовали данные 2017 г. На рис. 1 представлены графики индекса солнечных вспышек Fi за 2017 г. и значения прогноза на следующий день. Можно отметить что, несмотря на некоторое отличие в амплитуде прогнозируемые значения имеют ход во времени близкий к реаль-

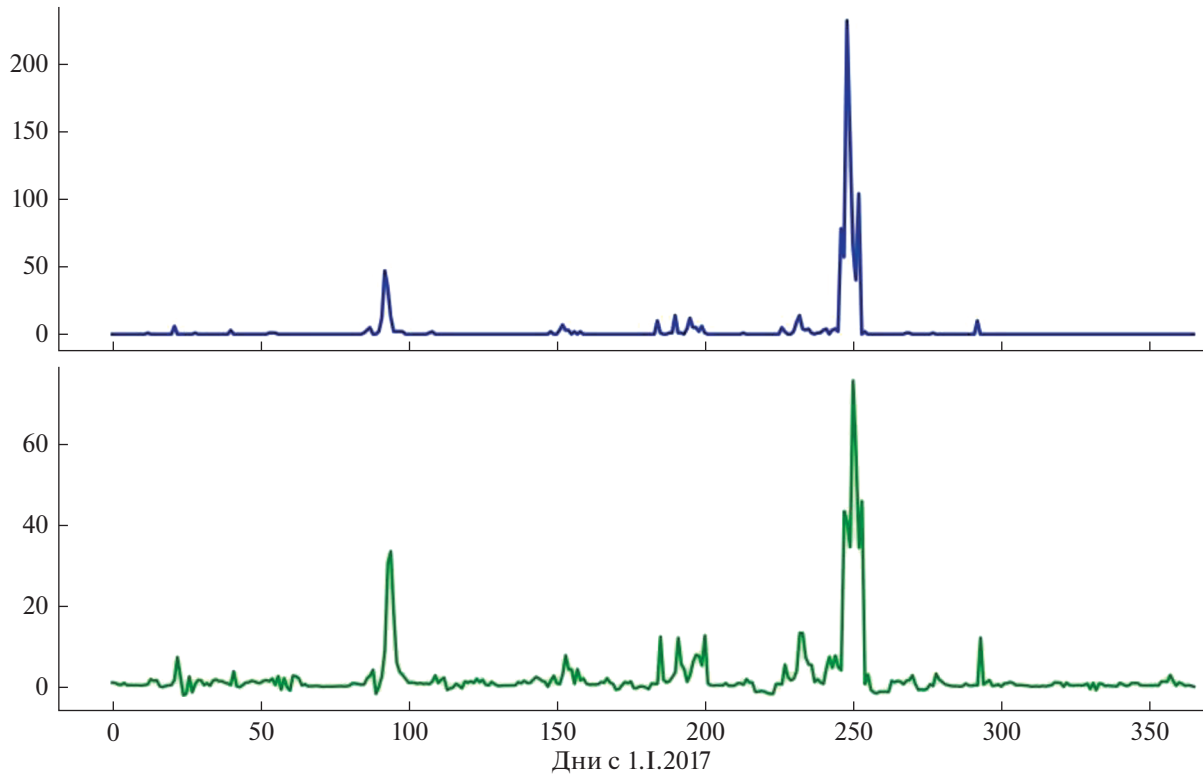


Рис. 1. Индекс солнечных вспышек F_i в 2017 г. (вверху) и прогноз индекса солнечных вспышек, рассчитанный для следующего дня (внизу).

ному индексу солнечных вспышек. В табл. 1 представлена значимость различных индексов в прогнозе солнечных вспышек. Наибольший вес имеют индексы рентгеновского фонового потока X_{bg} , потока радиоизлучения на волне 10.7 см и индекс солнечных вспышек за предыдущий день.

Для оценки степени достоверности прогнозов применяются различные критерии. В работе [30] предложены методы оценки точности, надежности прогнозирования параметров КП. Введены категориальные показатели эффективности прогноза, например, оптимальную вероятность и расчет частоты ложных срабатываний. Для оценки прогнозирования в работе [30] предлагаются использовать следующие параметры: A – количество попаданий, B – количество ложных срабатываний, C – количество пропущенных событий, D – количество правильных нулей, N – общее количество испытаний.

Из этих параметров составляется матрица ошибок (табл. 2), которая используется для расчета статистических показателей качества прогноза: чувствительность прогноза (*true positive rate*, TPR) $TPR = A/(A + C)$; частота ложных срабатываний (*false positive rate*, FPR) $FPR = B/(A + B)$; точность (*accuracy*) $= (A + D)/N$.

Матрица ошибок позволяет рассчитать еще одну метрику, часто применяемую для оценки ка-

чества прогноза – *Heidke Skill Score (HSS)* [31]. Эта метрика выражает то, насколько далек прогноз от случайного прогноза. Математическое выражение для метрики $HSS = 2(AD - BC)/((A + C)(C + D) + (A + D)(B + D))$. Показатель HSS может варьироваться от $-\infty$ до 1. Нулевое и отрицательные значения означают, что прогноз не лучше случайного прогноза. Для абсолютно точного прогноза показатель равен 1.

Для сильных вспышек мощностью класса M и более наша модель дала следующие параметры:

Таблица 1. Значимость различных индексов в прогнозе солнечных вспышек

№	Featrure	Importance
1	<i>X-Ray Bkgr Flux Value, X_{bg}</i>	0.227
2	<i>Radio 10.7 см, R_{10}</i>	0.169
3	<i>Flare index, F_i</i>	0.150
4	<i>Sunspot Number, R</i>	0.122
5	<i>Sunspot Area, S</i>	0.0900
6	Число вспышек класса C , N_C	0.086
7	<i>Flux Ratio, R_{10}/R_5</i>	0.086
8	<i>Radio 5 см, R_5</i>	0.065
9	Число вспышек класса M , N_M	0.005
10	Число вспышек класса X , N_X	0.00019

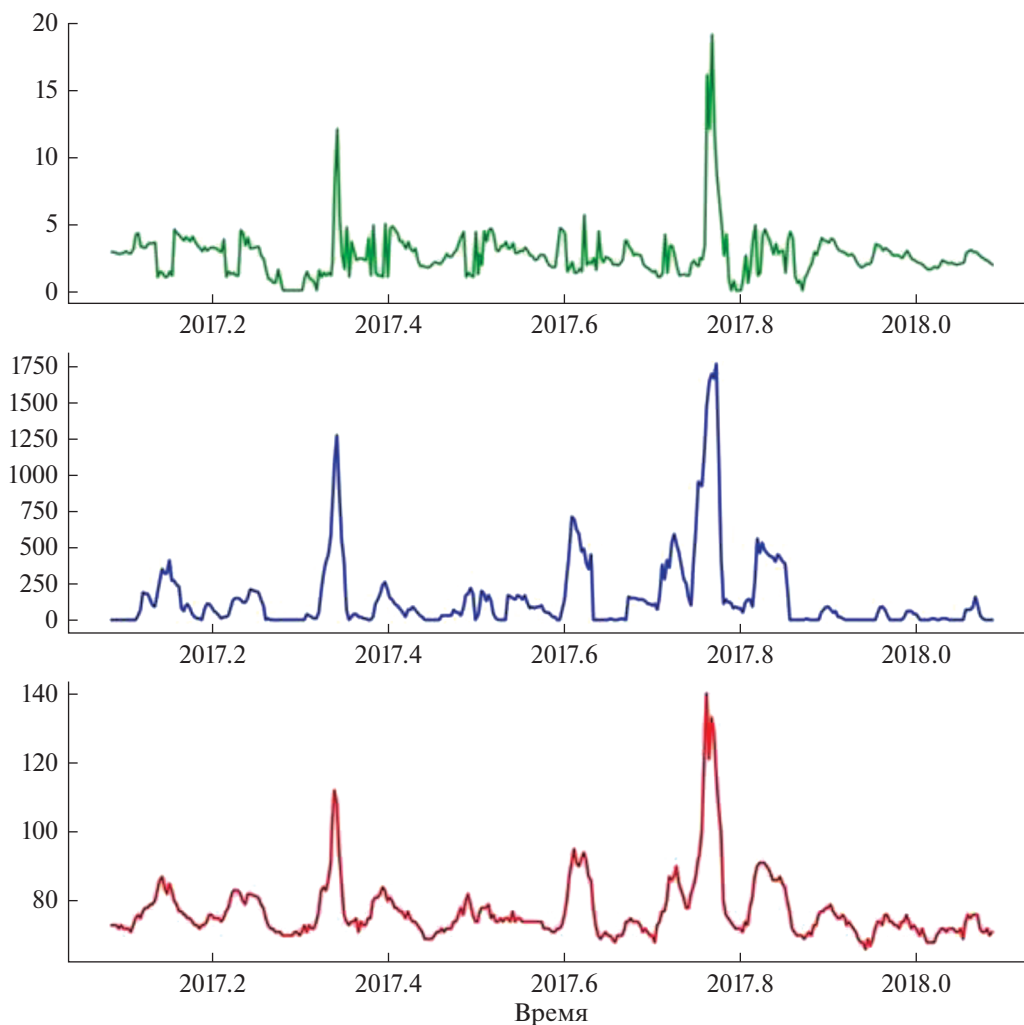


Рис. 2. Индекс изменения ежедневных значений индекса Xbg (вверху); индекс площади пятен S и потока радиоизлучения R_{10} .

$TPR = 0.972$; $HSS = 0.762$; $FPR = 0.552$; $accuracy = 0.773$. Корреляция между 2-мя графиками на рис. 1 составила $r = 0.947$. Такие показатели свидетельствуют о достаточно высокой эффективности предложенного прогноза.

ПРОГНОЗИРОВАНИЕ ПОТОКОВ ФОНОВОГО РЕНТГЕНОВСКОГО ИЗЛУЧЕНИЯ

Другим важным параметром в прогнозировании космической погоды является поток фонового рентгеновского излучения Xbg . Этот параметр связан с прогнозированием состояния ионосферы и широко применяется для оценки условий распространения радиоволн. Аналогично прогнозированию солнечных вспышек была разработана модель на основе сверточной нейронной сети DSD . В качестве прогнозируемого параметра мы использовали значения фонового потока рентгенов-

ского излучения Xbg в диапазоне $1-8 \text{ \AA}$, взятых из базы данных ftp://ftp.swpc.noaa.gov/pub/indices/old_indices/. В качестве исходных данных для прогноза мы использовали значения, доступные только с наземных телескопов: S , R_{10} , R_5 , S_{\max}^{gr} , $S_{\text{spot}}^{\text{max}}$. Здесь S_{\max}^{gr} и $S_{\text{spot}}^{\text{max}}$ – площадь максимальной группы пятен и площадь наибольшего пятна соответственно. На рис. 2 представлены графики

Таблица 2. Матрица ошибок для анализа качества прогнозов согласно [30]

Предсказанное значение	Yes	No
Yes	A	B
No	C	D

Где: A – количество попаданий, B – количество ложных срабатываний, C – количество пропущенных событий, D – количество правильных нулей, N – общее количество испытаний.

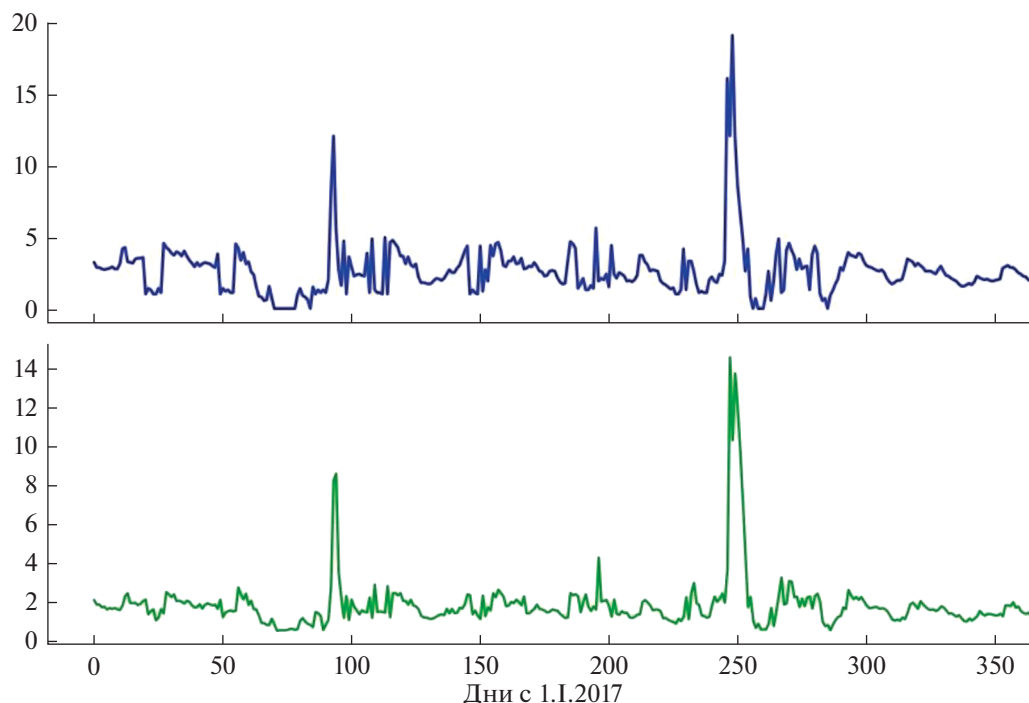


Рис. 3. Индекс фонового потока рентгеновского излучения Xbg в 2017 г. (вверху) и прогностический индекс Xbg , рассчитанный для следующего дня (внизу).

изменения ежедневных значений индекса Xbg , а также общей площади пятен S и потока радиоизлучения R_{10} для сравнения. Коэффициент корреляции между индексами Xbg и S составил $r = 0.57$, между Xbg и R_{10} , $r = 0.63$. Мы построили модель машинного обучения близкую по структуре к модели для прогнозирования солнечных вспышек. Индекс рентгеновского фонового потока Xbg мы рассчитывали по формуле $Xbg = 0.01k_A + 0.1k_B + 1k_C + 10k_M + 100k_X$, где k_A , k_B — значения потока для уровня потока A , B и так далее.

В данной модели мы использовали 2 скрытых слоя с количеством элементов 16 на входном слое и по 24 элемента в скрытых слоях. Коэффициент корреляции между реальным прогнозируемыми сигналами составило 0.69, то есть прогноз на следующий день имел лучшую корреляцию, чем корреляция между фоновым потоком и отдельными видами наблюдений для текущего дня (см. рис. 2 и 3).

ВЫВОДЫ

Мы разработали модель прогнозирования вспышек на основе методов машинного обучения, с использованием данных синоптических солнечных наблюдений, как наземных наблюдений, так и данных рентгеновского излучения спутников GOES. Используя обучающие и тестовые наборы данных, мы выполнили машинное обучение, чтобы предсказать индекс солнечных вспышек, которые происходят на следующий день.

Одной из целей этого анализа являлась задача определения, какие индексы солнечной активности являются наиболее важными в алгоритмах машинного обучения и наиболее подходят для модели прогнозирования вспышек. Оказалось для прогнозирования мощных солнечных вспышек на следующий день являются параметры значения фонового потока рентгеновского излучения Xbg , поток радиоизлучения на волне 10.7 см R_{10} и индекс солнечных вспышек текущего дня Fi (табл. 1). Вместе с тем наибольший вес параметров в прогнозе не превышает величины ~ 0.23 . Это значит, что для прогноза необходим набор параметров, которые характеризуют активность на различных высотах солнечной атмосферы.

Работа выполнена при частичной поддержке Российским фондом фундаментальных исследований, проекты № 18-52-34004 и 18-02-00098, и Программой ОФН РАН 12. Е.И. благодарит РФФ проект 20-72-00106.

СПИСОК ЛИТЕРАТУРЫ

1. Priest E.R., Forbes T.G. The magnetic nature of solar flares // *Astronomy and Astrophysics Review*. 2002. V. 10. P. 313.
2. Shibata K., Magara T. Solar Flares: Magnetohydrodynamic Processes // *Living Reviews in Solar Physics*. 2011. V. 8. № 6.
3. McIntosh P.S. The Classification of Sunspot Groups // *Sol. Phys.* 1990. V. 125. P. 251.

4. *Sammis, I., Tang, F., Zirin, H.* The Dependence of Large Flare Occurrence on the Magnetic Structure of Sunspots. // *Astrophysical J.* 2000. V. 540. P. 583.
5. *Li R., Cui Y., He H., Wang H.* Application of support vector machine combined with K-nearest neighbors in solar flare and solar proton events forecasting // *Advances in Space Research.* 2008. V. 42. P. 1469.
6. *Colak T., Qahwaji R.* Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares // *Space Weather.* 2009. V. 7. S06001.
7. *Bloomfield D.S., Higgins P.A., McAteer R.T.J., Gallagher P.T.* Toward Reliable Benchmarking of Solar Flare Forecasting Methods // *Astrophysical J. Letters.* 2012. V. 747. L41.
8. *Barnes G., Leka K.D., Schrijver C.J. et al.* A Comparison of Flare Forecasting Methods. I. Results from the “All-Clear” Workshop // *Astrophysical J. Letters.* 2016. V. 829. P. 89.
9. *Lee K., Moon Y.-J., Lee J.-Y. et al.* Solar Flare Occurrence Rate and Probability in Terms of the Sunspot Classification Supplemented with Sunspot Area and Its Changes // *Sol. Phys.* 2012. V. 281. P. 639.
10. *Zirin H.* *Astrophysics of the Sun.* Cambridge: Cambridge Univ. Press, 1988.
11. *Yu D., Huang X., Wang H., Cui Y.* Short-Term Solar Flare Prediction Using a Sequential Supervised Learning Method // *Sol. Phys.* 2009. V. 255. P. 91.
12. *Qahwaji R., Colak T.* Automatic Short-Term Solar Flare Prediction Using Machine Learning and Sunspot Associations. // *Sol. Phys.* 2007. V. 241. P. 195.
13. *Georgoulis M.K., Rust D.M.* Quantitative forecasting of major solar flares // *Astrophys. J.* 2007. V. 661. L109–L112.
14. *Schrijver C.J.*, A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting // *Astrophys. J.* 2007. V. 655. L117–L120.
15. *Falconer D., Barghouty A.F., Khazanov I., Moore R.* A tool for empirical forecasting of major flares, coronal mass ejections, and solar particle events from a proxy of active-region free magnetic energy // *Space Weather.* 2011. V. 9. S04003.
16. *Song H., Tan C., Jing J. et al.* Statistical assessment of photospheric magnetic features in imminent solar flare predictions // *Sol. Phys.* 2009. V. 254. P. 101–125.
17. *Mason J.P., Hoeksema J.T.* Testing automated solar flare forecasting with 13 years of Michelson Doppler Imager magnetograms // *Astrophys. J.* 2010. V. 723. P. 634–640.
18. *Falconer D.A., Moore R.L., Barghouty A.F., Khazanov I.* MAG4 versus alternative techniques for forecasting active region flare productivity // *Space Weather.* 2014. V. 12. P. 306–317.
19. *Boucheron L.E., Al-Ghraibah A., McAteer R.T.J.* Prediction of Solar Flare Size and Time-to-Flare Using Support Vector Machine Regression // *Astrophysical J.* 2015. V. 812. P. 51.
20. *Ahmed O.W., Qahwaji R., Colak T. et al.* Solar flare prediction using advanced feature extraction, machine learning, and feature selection // *Sol. Phys.* 2013. V. 283. P. 157–175.
21. *Higgins P.A., Gallagher P.T., McAteer R.T.J., Bloomfield D.S.* Solar magnetic feature detection and tracking for space weather monitoring // *Advances in Space Research.* 2011. V. 47. P. 2105.
22. *Nishizuka N., Sugiura K., Kubo Y. et al.* Solar Flare Prediction Model with Three Machine-learning Algorithms using Ultraviolet Brightening and Vector Magnetograms // *Astrophysical J.* 2017. V. 835. P. 156.
23. *Yu D., Huang X., Hu Q. et al.* Short-term Solar Flare Level Prediction Using a Bayesian Network Approach // *Astrophysical J.* 2010. V. 709. P. 321.
24. *Al-Ghraibah A., Boucheron L.E., McAteer R.T.J.* Prediction of Solar Flare Size and Time-to-Flare Using Support Vector Machine Regression // *Astronomy & Astrophysics.* 2015. V. 579. A64.
25. *Boucheron L.E., Al-Ghraibah A., McAteer R.T.J.* Prediction of Solar Flare Size and Time-to-Flare Using Support Vector Machine Regression // *Astrophysical J.* 2015. V. 812. P. 51.
26. *Guerra J.A., Pulkkinen A., Uritsky V.M.* Ensemble forecasting of major solar flares: First results // *Space Weather.* 2015. V. 13. P. 626.
27. *Abramenko V.I.* Multifractal analysis of solar magnetograms // *Sol. Phys.* 2005. V. 228. P. 29–42.
28. *Han S., Pool J., Narang S. et al.* DSD: Dense-Sparse-Dense Training for Deep Neural Networks // *International Conference on Learning Representations (ICLR).* 2008.
29. *Breiman L.* *Random Forests* // *Machine Learning.* 2001. V. 45. P. 5–32.
30. *Balch C.* Updated verification of the Space Weather Prediction Center’s solar energetic particle prediction model // *Space Weather.* 2008. V. 6. S01001.
31. *Heidke P.* Berechnung des Erfolges und der Gute. der Windstarkevorhersagen im Sturmwarnungsdienst // *Geogr. Ann.* 1926. V. 8. P. 301–349.