

Article

Detecting Blood Methylation Signatures in Response to Childhood Cancer Radiotherapy via Machine Learning Methods

Zhandong Li ^{1,†}, Wei Guo ^{2,†}, Shijian Ding ^{3,†}, Kaiyan Feng ⁴, Lin Lu ^{5,*}, Tao Huang ^{6,7,*}  and Yudong Cai ^{3,*} 

¹ College of Food Engineering, Jilin Engineering Normal University, Changchun 130052, China; lizd591@jlenu.edu.cn

² Key Laboratory of Stem Cell Biology, Shanghai Jiao Tong University School of Medicine (SJTUSM) & Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai 200025, China; gw_1992@alumni.sjtu.edu.cn

³ School of Life Sciences, Shanghai University, Shanghai 200444, China; dingshijian@shu.edu.cn

⁴ Department of Computer Science, Guangdong AIB Polytechnic College, Guangzhou 510507, China; kyfeng@gdaib.edu.cn

⁵ Department of Radiology, Columbia University Medical Center, New York, NY 10032, USA

⁶ Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

⁷ CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

* Correspondence: ll2860@cumc.columbia.edu (L.L.); tohuangtao@126.com (T.H.); cai_yud@126.com or caiyudong@staff.shu.edu.cn (Y.C.); Tel.: +86-21-54923269 (T.H.); +86-21-66136132 (Y.C.)

† These authors contributed equally to this work.



Citation: Li, Z.; Guo, W.; Ding, S.; Feng, K.; Lu, L.; Huang, T.; Cai, Y. Detecting Blood Methylation Signatures in Response to Childhood Cancer Radiotherapy via Machine Learning Methods. *Biology* **2022**, *11*, 607. <https://doi.org/10.3390/biology11040607>

Academic Editor: Holger Fröhlich

Received: 25 February 2022

Accepted: 14 April 2022

Published: 15 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Radiotherapy for cancer patients can cause abnormal DNA methylation. We developed a computational workflow that can identify crucial methylation alterations related to treatment exposure in childhood cancer survivors.

Abstract: Radiotherapy is a helpful treatment for cancer, but it can also potentially cause changes in many molecules, resulting in adverse effects. Among these changes, the occurrence of abnormal DNA methylation patterns has alarmed scientists. To explore the influence of region-specific radiotherapy on blood DNA methylation, we designed a computational workflow by using machine learning methods that can identify crucial methylation alterations related to treatment exposure. Irrelevant methylation features from the DNA methylation profiles of 2052 childhood cancer survivors were excluded via the Boruta method, and the remaining features were ranked using the minimum redundancy maximum relevance method to generate feature lists. These feature lists were then fed into the incremental feature selection method, which uses a combination of deep forest, k-nearest neighbor, random forest, and decision tree to find the most important methylation signatures and build the best classifiers and classification rules. Several methylation signatures and rules have been discovered and confirmed, allowing for a better understanding of methylation patterns in response to different treatment exposures.

Keywords: methylation; childhood cancer radiotherapy; machine learning method; feature selection; rule learning

1. Introduction

Radiotherapy (RT) has been an important and effective anticancer treatment for over a century. Approximately 70% of all patients with cancer are treated via RT alone or in combination with other treatment approaches [1]. The application of RT in cancer treatment has largely improved the short-term survival of patients [2]. High doses of radiation can kill cancer cells and shrink tumors as electrically charged particles pass

through tumor cells. DNA double-strand breaks are the classical outcomes induced by RT that can effectively arrest cell growth and induce cell death in tumor cells [3]. However, RT itself can also cause DNA damage in normal tissues and result in long-term morbidity and mortality. Many underlying molecular alterations induced by RT can cause long-term adverse health outcomes. Patients with cancer who had received RT reportedly suffered from sterile inflammation, premature senescence, and cardiometabolic diseases during long-term outcomes [4]. Therefore, biomarkers that indicate the potential risk of long-term outcomes after RT exposure must be identified.

Various cancer treatments, such as drug treatment, RT, or the combination of anticancer treatments, can potentially impact the methylation status of the genome, subsequently leading to gene regulation alterations and aberrant phenotypes [5]. Unlike genetics, which is relatively static and exerts a direct effect on gene encoding, methylation modification is thought to be plastic and can be modified in response to environmental stimulations [6]. Radioactive exposure might cause the acquisition of soma-wide alterations in DNA methylation. Accumulating evidence supports the idea that DNA methylation abnormalities are closely associated with a diverse group of diseases [7]. The direct effects of radiation on DNA methylation had been reported as early as 1989 by Kalinich et al. [8]. They observed a decrease in 5-methylcytosine after γ radiation irradiation in cell lines in vitro. Pogribny et al. [9] also found different DNA methylation patterns at various doses of X-ray exposure. A hypothesis suggests that the alteration of DNA methylation may reflect biological responses to radiation that will lead to specific sensitivity to RT [10]. Although the association between DNA methylation alteration and health outcomes has been widely reported, the underlying biological mechanisms by which radioactive exposure affects methylation modifications are still incompletely understood. Moreover, the methylation patterns induced by region-specific RT require further research. Therefore, we focused on the key DNA methylation alterations associated with specific cancer RT and the functional role of such methylations in health outcomes.

In this study, we computationally analyzed the methylation profiles of patients who underwent RT. A recent publication obtained the methylation data from 2052 cancer survivors by using DNA methylation microarray [11]. On the basis of these public data, we conducted a machine learning analysis to identify the key methylation sites that may be relevant to RT exposure and its long-term adverse health outcomes. Furthermore, we divided this cohort into four categories according to the types of treatment exposures, namely, abdominal RT, brain RT, chest RT, and pelvic RT. Subsequently, we applied the minimum redundancy maximum relevance (mRMR) and incremental feature selection (IFS) methods to identify the most relevant methylation sites for predicting each type of RT, and then we constructed decision rules for the quantitative description of the relationship between the methylation sites and RT. Overall, our study sheds light on the potential methylation modifications in response to region-specific cancer RT.

2. Materials and Methods

2.1. Datasets

The methylation datasets of childhood cancer survivors were obtained from the Gene Expression Omnibus database with the accession number GSE169156 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE169156> (accessed on 6 April 2021)) [11]. The datasets included the blood DNA methylation profiles of patients who underwent abdominal RT, brain RT, chest RT, and pelvic RT. Table 1 shows the number of positive and negative samples for each RT. Each sample had the methylation levels of 865,892 sites measured with Illumina Infinium HumanMethylation850 BeadChip which included 866,895 probes. Some probes had too many missing values and were excluded from further analysis. The beta values of 865,892 methylation sites were analyzed. The beta values ranged from 0 to 1. A high beta value meant methylated while a low beta value meant unmethylated. All the data descriptions can be found at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE169156> (accessed on 6 April 2021) and the annotation infor-

mation of Illumina Infinium HumanMethylation850 BeadChip can be found at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL23976> (accessed on 6 April 2021).

Table 1. Sample size of patients treated with different radiotherapy (RT).

Dataset	Positive Sample	Negative Sample	Total
Abdominal RT	412	1640	2052
Brain RT	629	1423	2052
Chest RT	577	1475	2052
Pelvic RT	352	1700	2052

2.2. Boruta Feature Selection

Directly analyzing all methylation features in a dataset is difficult and time-consuming owing to the vast number of features in the dataset. In this study, we applied the Boruta feature selection method to remove irrelevant features [12]. The shadow features in Boruta were created by shuffling the original features, and then the feature matrix was created by connecting the original features and the shadow features were trained using random forest (RF). Finally, the feature importance score of each shadow feature was utilized as a reference to pick the feature set from the original features most closely related to the category variables. The methylation profiles were analyzed using the Boruta program (https://github.com/scikit-learn-contrib/boruta_py (accessed on 14 September 2020)), and the default parameters were adopted to run the program for convenience.

2.3. mRMR

mRMR is a powerful method in feature selection and has been widely applied in the field of biomedical research [13]. It evaluates the importance of features on the basis of mutual information (MI), which is defined as follows:

$$I(X, Y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy, \quad (1)$$

where $p(x, y)$ is the joint probability density function of x and y ; and $p(x)$ and $p(y)$ are the marginal probability density functions of x and y , respectively. A high MI value indicates a great correlation between X and Y . Suppose S represents the set of features that have been selected, let us define the following optimization equations:

$$\max D(S, c), \quad D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c). \quad (2)$$

However, the features of S may have redundancies. The redundancy of set S is as follows:

$$\min R(S), \quad R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j), \quad (3)$$

The objective of mRMR is to select the set S with the maximum relevance and the minimum redundancy, which is defined as follows:

$$\max \Phi(D, R), \quad \Phi = D - R. \quad (4)$$

Therefore, an increase in D and a decrease in R both contribute to an increase in the objective function. Given that there are already $m - 1$ features in S , then the m -th feature is selected from the remaining features to maximize $\Phi(D, R)$. In the end, mRMR outputs a ranked feature list of m features. In the present study, the mRMR program was obtained from <http://www.home.penglab.com/proj/mRMR/> (accessed on 2 May 2018) and run with the default parameters.

2.4. IFS

IFS can determine the best number of features by using machine learning algorithms, such as RF, in a ranked feature list [14]. IFS can construct a succession of feature subsets on the basis of a given step interval s (i.e., 1) for a feature list F generated by mRMR. For example, the first feature subset F_1 contains the top $1 \times s$ features, whereas F_2 contains the top $2 \times s$ features, and so on. For each candidate feature subset F_i , a machine learning model will be trained on the samples comprising this feature subset. Using the 10-fold cross-validation [15] and synthetic minority oversampling technique (SMOTE) procedures, the evaluation metrics indicating the model's performance, such as Matthews correlation coefficient (MCC), are obtained. Finally, the IFS curves are produced using the number of features as the x-coordinate and an evaluation metric as the y-coordinate. The highest point of the curves can be used to identify the best feature subset.

2.5. Classification Algorithms

In the IFS process step, four classification algorithms are used to build classifiers, namely, DF [16], kNN [17], RF [18], and DT [19], which are described in detail below. These algorithms have been widely used in tackling various medical problems [20–30].

2.5.1. DF

DF combines numerous ensemble-based methods, such as RFs and stacking, to build a cascade structure that resembles a multilayer neural network, but each layer contains RFs instead of neurons. Each layer accepts the feature information processed by the previous layer and outputs the result to the next layer in this architecture. A multigranularity scanning method can increase the representation learning capabilities of the input with large dimensionality. The DF program was downloaded from <https://github.com/LAMDA-NJU/Deep-Forest> (accessed on 15 November 2020) and run with the default parameters.

2.5.2. kNN

kNN is a basic supervised learning algorithm. The key idea of this algorithm is to calculate the distance (e.g., Euclidean distance) between a new instance and each training sample and then find the first k -nearest samples and determine the category of the new instance.

2.5.3. RF

RF is an ensemble learning method that improves a model's prediction ability by integrating a number of DTs. Each DT is trained using some randomly selected samples and features from the original dataset. For a test sample, RF integrates the decisions of each DT to arrive at the final decision by majority voting.

2.5.4. DT

DT is a white-box model that gives interpretable decision rules, unlike the three machine learning methods discussed above. To divide occurrences and features, it creates a classification or a regression model on the basis of the IF-THEN structure. In this study, the Scikit-learn package was applied to execute kNN, RF, and DT by using the default parameters.

2.6. SMOTE

In this study, SMOTE was used to generate the sample data of minority classes because four methylation profiles are highly uneven [31]. According to the principle of kNN, this method calculates the distances between a sample and other samples in the minority class and then selects multiple samples, including the sample itself and some of its neighbors, to generate a new sample linearly. SMOTE was employed to balance the training set so that the number of samples from different classes was equal when evaluating the performance

of a classification model with 10-fold cross-validation. The SMOTE program with default parameters from the imbalance-learn package was utilized for this analysis.

2.7. Performance Measurement

In the process of 10-fold cross-validation, classification accuracy (ACC), specificity (SP), sensitivity (SN), and MCC [32–35] were used as evaluation metrics. These metrics are calculated as follows:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

$$SP = \frac{TN}{TN + FP} \quad (6)$$

$$SN = \frac{TP}{TP + FN} \quad (7)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

where TP , TN , FP , and FN represent true positive, true negative, false positive, and false negative, respectively.

3. Results

In the present study, we proposed a computational workflow for analyzing the DNA methylation profiles of patients who underwent RT. Various feature selection methods and classification algorithms were used. Figure 1 depicts the entire study's analysis flow, and the findings of this study are also presented.

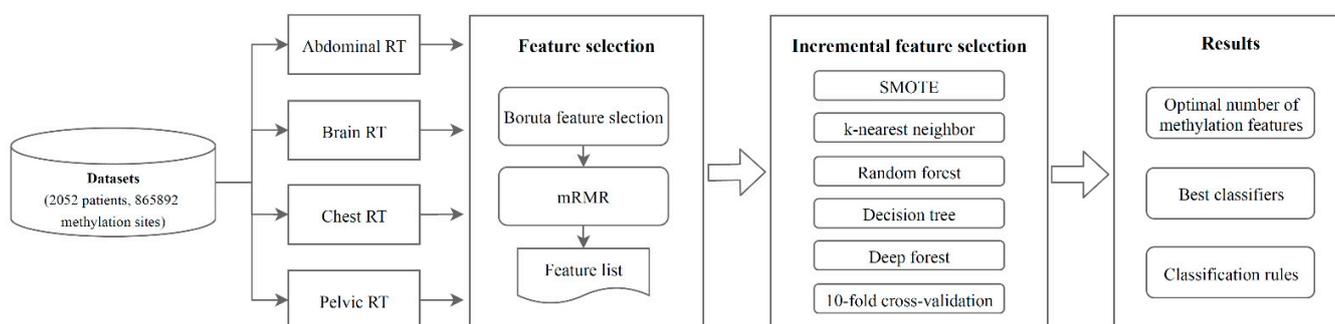


Figure 1. Computational workflow of this study. The methylation dataset was acquired in four sections from a public database: abdominal RT, brain RT, chest RT, and pelvic RT. The methylation features in each methylation profile were filtered and ranked using the Boruta feature selection and mRMR methods. The IFS method used the resulting feature list to identify the optimal number of features and develop the best classifiers and classification rules by combining SMOTE and classification algorithms.

3.1. Results of Feature Selection of the Methylation Datasets via the Boruta and mRMR Methods

Each original dataset contained 865,892 methylation sites, and the computational complexity of direct analysis was enormous. To address this issue, we initially employed the Boruta method to filter the features from the DNA methylation profiles of four tissues from patients who underwent RT. As a consequence, the number of retained features for abdominal RT, brain RT, chest RT, and pelvic RT was 766, 155, 972, and 257, respectively. The mRMR method was then utilized to construct four ranked feature lists according to the mRMR criterion, as shown in Table S1.

3.2. Results of IFS Method with Classification Algorithms

The feature lists sorted by the mRMR method were fed into the IFS method with four classification algorithms to determine the best number of features. When the step

size was set to 1, the first feature subset produced by IFS was the first feature in the list, the second feature subset was made up of the top two features, and so on. For example, the abdominal RT dataset yielded 766 feature subsets. Subsequently, four classification algorithms, namely, DF, kNN, RF, and DT, were adopted to build classifiers by using the sample data represented by these feature subsets and evaluation metrics were obtained. The performance of these classifiers with different feature subsets in different methylation datasets is provided in Tables S2–S5. The IFS curves were plotted with MCC as the vertical coordinate and the number of features as the horizontal coordinate, as shown in Figures 2–5.

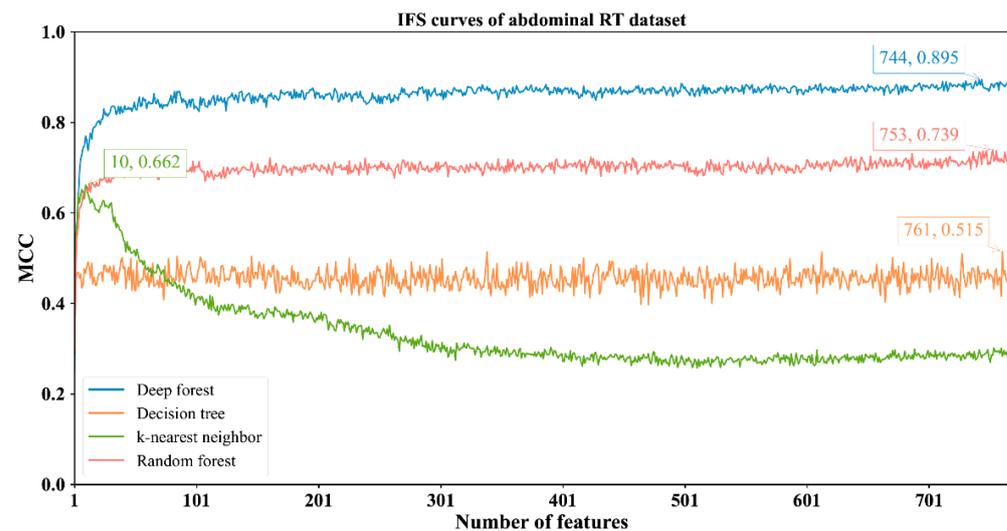


Figure 2. IFS curves with different classifiers on the different numbers of methylation features for abdominal RT methylation dataset. DF achieves the highest MCC value of 0.895 when the top 744 features are used.

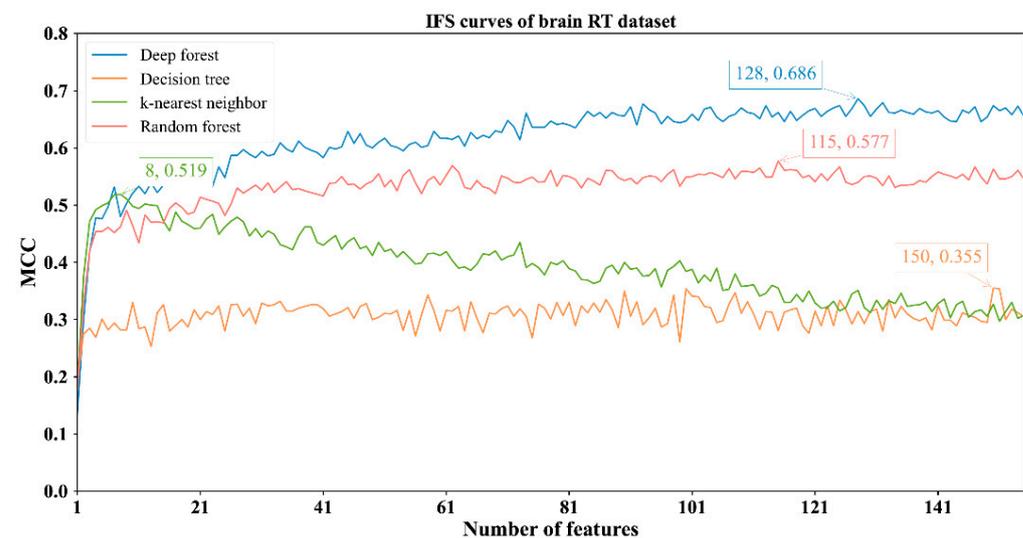


Figure 3. IFS curves with different classifiers on the different numbers of methylation features for brain RT methylation dataset. DF attains the highest MCC value of 0.686 when the top 128 features are utilized.

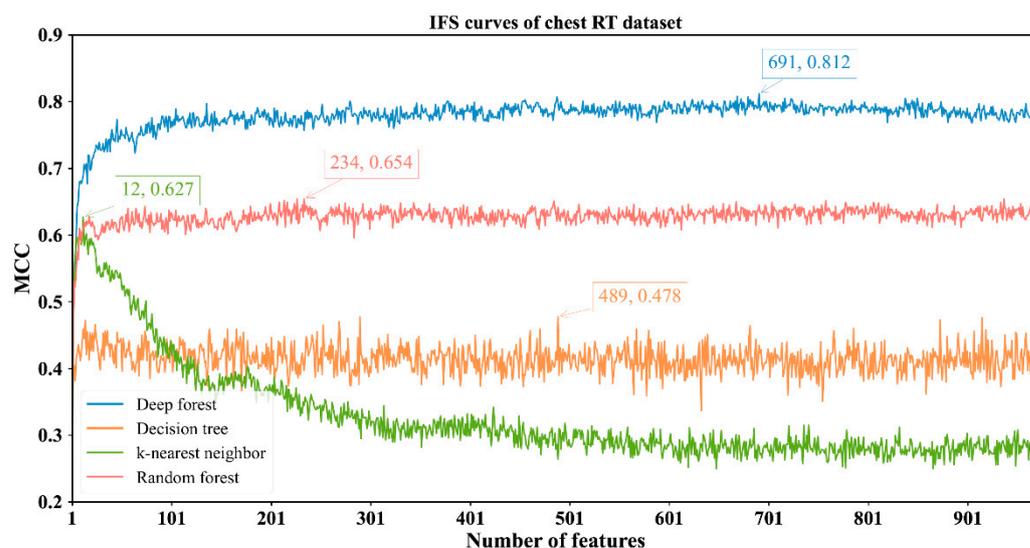


Figure 4. IFS curves with different classifiers on the different numbers of methylation features for chest RT methylation dataset. DF reaches the highest MCC value of 0.812 when the top 691 features are adopted.

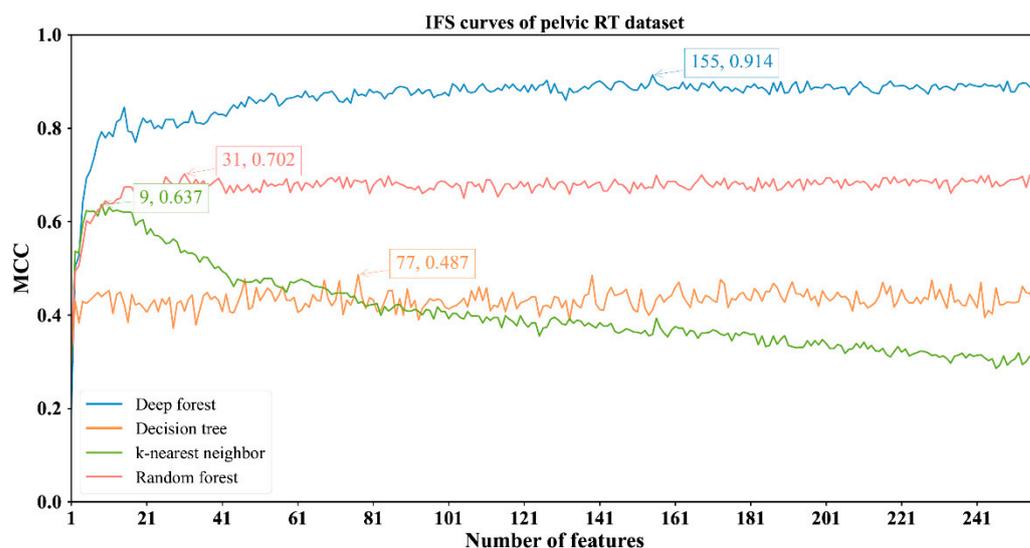


Figure 5. IFS curves with different classifiers on the different numbers of methylation features for pelvic RT methylation dataset. DF yields the highest MCC value of 0.914 when the top 155 features are employed.

For the abdominal RT dataset, it can be observed from Figure 2 that four classification algorithms (DF, kNN, RF and DT) yielded the highest MCC of 0.895, 0.662, 0.739 and 0.515, respectively. These MCCs were obtained by using top 744, 10, 753, 761 features. Other measurements of these classification algorithms under corresponding features are listed in Table 2. Evidently, DF with the top 744 features provided the best performance. The MCC was at least 15% higher than the other highest MCCs yielded by the other three classification algorithms, suggesting the superiority of DF for identifying samples with abdominal RT.

Table 2. Detailed performance of different classifiers on four methylation datasets.

Dataset	Classifiers	Number of Features	Accuracy	Sensitivity	Specificity	MCC
Abdominal RT	DF	744	0.966	0.910	0.980	0.895
	kNN	10	0.846	0.971	0.814	0.662
	RF	753	0.903	0.913	0.900	0.739
	DT	761	0.791	0.825	0.783	0.515
Brain RT	DF	128	0.869	0.736	0.928	0.686
	kNN	8	0.749	0.863	0.699	0.519
	RF	115	0.811	0.765	0.832	0.577
	DT	150	0.690	0.693	0.688	0.355
Chest RT	DF	691	0.925	0.828	0.963	0.812
	kNN	12	0.804	0.945	0.749	0.627
	RF	234	0.851	0.823	0.862	0.654
	DT	489	0.762	0.747	0.768	0.478
Pelvic RT	DF	155	0.976	0.923	0.986	0.914
	kNN	9	0.841	0.977	0.813	0.637
	RF	31	0.896	0.906	0.894	0.702
	DT	77	0.798	0.795	0.798	0.487

For the other three datasets, DF still provided the highest MCC. In detail, DF with the top 128 features yielded the MCC of 0.686 on the brain RT dataset (Figure 3); DF with the top 691 features produced the MCC of 0.812 on the chest RT dataset (Figure 4); DF with the top 155 features generated the MCC of 0.914 on the Pelvic RT dataset (Figure 5). The ACC values of these DF classifiers were 0.869, 0.925, and 0.976 (Table 2), respectively. These values were all higher than those of the other three classification algorithms (Table 2). MCC was at least 10% higher and ACC was 5% higher, suggesting DF can capture more essential information in these datasets, thereby building more efficient classifiers.

From Figures 2–5 and Table 2, some interesting phenomena can be observed. First, the performance of DF, kNN, RF and DT was uniform on four datasets. DF gave the best performance, followed by RF, kNN and DT. This result almost conformed to our general cognition. DF can be deemed as a generalized version of RF. Thus, it is generally more powerful than RF. kNN, in fact, is not a pure machine learning algorithm because it does not contain the training procedures. In most cases, it is weaker than RF. DT, as a rule-learning algorithm, cannot always provide high performance. Thus, its performance was the lowest in this study. However, its classification procedures are completely open, providing more clues to uncovering essential information behind the dataset. Second, the best kNN classifier adopted much fewer features than other three best classifiers on all datasets. kNN used about ten features to generate the highest MCC, whereas the other three algorithms need tens of, or even hundreds of, features to achieve the highest MCC. In the feature list yielded by the mRMR method, features with high ranks had a higher relationship with class labels. With a small number of top features in the list, kNN can easily distinguish positive and negative samples using a simple way (distance between samples). The other three algorithms adopted a much more complicated scheme to train the classifiers, these features were too few to build the optimum classifiers. However, when more and more features were added, more noises were included. As kNN does not contain the training procedure, it cannot identify interference information and exclude it, thereby influencing its performance. For the other three classification algorithms, their training procedures can help them extract useful information and build more powerful classifiers.

3.3. Classification Rules Extracted by the Optimal DT Classifiers

DF performed well in each methylation dataset. However, it is a black-box model that cannot provide quantitative rules. To extract the decision rules, we used the top 761, 150,

489, and 77 features from the abdominal RT, brain RT, chest RT, and pelvic RT datasets, respectively, to build the best DT classifiers. The expression rules obtained by the optimal DT classifier for each dataset are provided in Table S6. The abdominal RT, brain RT, chest RT, and pelvic RT datasets had 151, 239, 166, and 183 rules each. The number of rules for the positive and negative classes on each dataset is listed in Table 3.

Table 3. Total number of rules and the number of rules with each category generated by the optimal DT classifier in the four datasets.

Dataset	Number of Rules	Number of Rules for Positive Class	Number of Rules for Negative Class
Abdominal RT	151	87	64
Brain RT	239	132	107
Chest RT	166	93	73
Pelvic RT	183	99	84

4. Discussion

This study demonstrated that several optimal classifiers can recognize risk conditions, such as region-specific RT exposure, with relatively high ACC values on the basis of methylation profiles. In detail, we treated the methylation level of each site as the feature and identified the most relevant features through the Boruta and mRMR methods. The crucial DNA methylations that indicated abdominal RT, brain RT, chest RT, and pelvic-RT were individually estimated. We applied four different algorithms, namely, DF, kNN, RF, and DT, to construct the classifiers. DF was shown to have the best performance in the classification. Via feature selection, we identified 744 DNA methylation sites that were highly predictive of abdominal RT treatment. Moreover, we found that 128 crucial DNA methylation sites were associated with brain RT treatment. Furthermore, we determined that 691 DNA methylation sites were related to chest RT treatment. Finally, we recognized 155 key DNA methylation sites linked to pelvic RT treatment. We noticed that many methylation sites identified by our analysis had been reported to be significantly associated with RT treatment via epigenome-wide association study (EWAS) method by Song et al. [11], confirming the reliability of these feature selection methods. In detail, we compared the most relevant features related to each RT exposure by our analysis to the significant methylation sites ($p < 9 \times 10^{-8}$) from a previous EWAS study. Among the 330 methylation sites reported to be significantly associated with abdominal RT in the EWAS study, there were 169 methylation sites identified as highly related to abdominal RT in this study. The previous EWAS study reported nine methylation sites significantly associated with brain RT, and two of them were identified as highly predictive of brain RT by feature selection. Next, among 303 methylation sites significantly associated with chest RT exposure, there were 157 identified in the present study with the most relevance to chest RT. A total of 248 methylation sites were reported to be associated with pelvic RT by EWAS analysis. Of these, 113 methylation sites were identified in feature selection of pelvic RT exposure. Taken together, almost half of the previously reported methylation sites associated with RT treatment were identified again using a distinct computational method.

Essentially, the EMAS method is a set of statistical analysis approaches. In this study, we adopted quite different computational methods, i.e., machine learning algorithms, to reanalyze the blood DNA methylation profiles. These algorithms can deeply mine hidden relationships behind the datasets, including relationships between features and class labels or among features, which cannot be discovered by general statistical analysis approaches. Furthermore, the training procedures of such algorithms can help us improve the performance of classifiers. Thus, we obtained a different ranking of the feature's relevance by comparison with the original EWAS study, showing an improved sensitivity and accuracy in identifying RT-related methylation modifications. The decision rules were built on the basis of the selected features, providing the criteria to indicate treatment exposures. To validate the relevance of these findings in distinguishing region-specific

RT exposure, functional characteristics of these methylation sites were gathered from the literature, which supported a potential association between function injury and each type of treatment exposure. For each category of RT exposure, we presented detailed descriptions of the functional role of methylation modification related to RT.

4.1. Key Methylation Alteration Related to Abdominal RT

The CpG site cg21585138, which is located on chr3:5064516 and is mapped to the *CISH* gene, was identified as one of the most relevant features for indicating abdominal RT. *CISH* is involved in the IL-2 signaling pathway, and it is reportedly associated with infectious diseases [36]. The loss of *CISH* contributes to hyperproliferative responses in acute myelogenous leukemia [37]. Additionally, the methylation status of cg21585138 was found to be influenced by smoking, suggesting a potential epigenetic alteration caused by chemical toxicity [38]. This evidence supported the contention that cg21585138 may serve as a methylation signature for the risk of adverse health conditions. Abdominal RT allegedly exerts a harmful effect on health outcomes, and the methylation alteration at cg21585138 may be the early event after RT.

Another key CpG site, cg03054277, was identified to be highly predictive of abdominal RT. This methylation site is located on chr1:228400217 and mapped to the *OBSCN* gene. The protein product encoded by *OBSCN* is related to various functions, including transferase activity and tyrosine kinase activity. *OBSCN* reportedly plays a role in mediating cardiomyocyte adhesion via PI3K/AKT/mTOR signaling [39]. An epigenome-wide association analysis revealed that the methylation status of cg03054277 is associated with age, implying that it may be a senescence-related signature [40]. The CpG site cg03054277 is also identified as a DNA methylation biomarker of alcohol consumption [41]. Given that alcohol intake is viewed to cause the accumulation of body lesions, the methylation status of cg03054277 may indicate an initial signal for chemical toxicity. Therefore, cg03054277 may also act as the signature for abdominal RT.

Among the decision rules for identifying abdominal RT exposure, the CpG site cg17730048 was hypermethylated to indicate abdominal RT. This methylation site is located on chr17:26577563 within the CpG island region. Notably, cg17730048 is also identified as one of the risk signals associated with aging, suggesting that a high methylation level of cg17730048 may represent the impaired functional condition of an individual [40]. Moreover, this CpG site has been linked to maternal smoking in pregnancy [42]. This finding supported the idea that hypermethylation of cg17730048 may indicate the risk for adverse health conditions, consistent with our analysis that the hypermethylation of cg17730048 can predict abdominal RT.

4.2. Key Methylation Alteration Related to Brain RT

The most relevant CpG site for brain RT we identified was cg08866213, which is located on chr3:192530777 and mapped to the *MB21D2* gene. Overexpressed *MB21D2* reportedly promotes a pro-oncogenic progression of head and neck cancer, and it also induces less sensitivity toward DNA-damaging agents, such as RT [43]. In addition, paclitaxel, a chemotherapy medication, allegedly results in the expression alteration of *MB21D2* [44]. These results suggested that the *MB21D2* gene may be a key target in response to DNA-damaging agents, such as RT and chemotherapy. Therefore, we argue that changes in the methylation of cg08866213 can serve as a biomarker for indicating brain RT.

We also identified the CpG site cg15393490 as another important feature to indicate brain RT. This methylation site is located on chr1:207996459 and belongs to the promoter region of miR-29c. TCGA data indicated different methylation levels of cg15393490 in breast tumor subtypes [44]. Notably, miR-29c has been shown to be involved in many types of diseases, including ischemic brain injury [45]. We inferred that the methylation status of cg15393490 may be a risk indicator for brain injury, for example, the damage caused by brain RT. Furthermore, the DNA methylation of cg15393490 is reportedly associated with liver diseases and cholesterol metabolism [46,47]. These findings implied that the

methylation status of cg15393490 is related to long-term adverse health conditions that may be caused by brain RT.

Among the rules identifying brain RT, we found that one CpG site, that is, cg18973101, was involved in several criteria. This CpG site is located on chr1:156251280 within the intergenic region between the *TMEM79* and *SMG5* genes. A high methylation level of cg18973101 is required to indicate brain RT. Several studies have discovered that the DNA methylation of cg18973101 is associated with long-term alcohol consumption [41,48]. The influence of alcohol consumption on the risk of disease is widely recognized, and DNA methylation may be one of the pathogenic mechanisms. We speculated that RT treatment may also exert a similar effect on DNA methylation alteration in cg18973101, making this key CpG site one of the signatures indicating brain RT exposure.

4.3. Key Methylation Alteration Related to Chest RT

We identified the CpG site cg01511232 as one of the most relevant features for predicting chest RT. This methylation site is located on chr4:155661929 and cannot be annotated to any known genes to date. Some pieces of evidence support the conjecture that the DNA methylation of cg01511232 is associated with immune regulation and the risk of HIV infection [49]. Age is also considered to be a factor that causes the methylation alteration of cg01511232 [50]. These results suggested that cg01511232 is a risk signal for adverse health outcomes.

The CpG site cg08601457, which is mapped to the *FYN* gene and located on chr6:112115117, was found to be strongly relevant to our classification. The related pathways of *FYN* are RET signaling and adherens junction. *FYN* is an important molecular marker in breast cancer that can serve as a predictor of early recurrence [51]. The heterogeneous expression of *FYN* is also reported to have prognostic implications in lymphoma. *FYN* overexpression promotes cell proliferation and cell migration in various types of cancers and mediates epithelial–mesenchymal transition [52,53]. Therefore, the epigenetic modification at cg08601457 likely has a substantial influence on cancer progress that may be induced by the radiation effect.

A relatively high methylation level of cg23752651 indicated chest RT in the decision rules. This site is mapped to the *TNFRSF1A* gene, which is related to tumor necrosis factor-activated receptor activity. The specific methylation pattern at cg23752651 has been found in pancreatic ductal adenocarcinoma [54]. The R92Q variant in the *TNFRSF1A* gene reportedly influences susceptibility and phenotype depending on the age at disease [55]. Moreover, *TNFRSF1A* purportedly may serve as a diagnostic and prognostic biomarker in gliomas [56]. The polymorphism of *TNFRSF1A* is regarded as a predictive factor for RT-induced oral mucositis [57]. Therefore, we speculate that the methylation status of cg23752651 can indicate chest RT.

4.4. Key Methylation Alteration Related to Pelvic RT

The methylation site cg20112376 was found to be highly relevant to the classification for pelvic RT. This CpG site is located on chr4:6118443 and mapped to the *JAKMIP1* gene. An epigenome-wide association study found that cg20112376 is associated with long-term exposure to noise and air pollution [58], suggesting that the methylation status of cg20112376 represents a damaging burden of disease. The protein encoded by *JAKMIP1* plays a role in regulating microtubule cytoskeleton rearrangements. Expression changes in *JAKMIP1* have been observed in the peripheral blood of patients undergoing RT [59]. In vitro experiments also demonstrated that radiation remarkably alters the expression of *JAKMIP1* [60]. These findings supported the reliability of our analysis that cg20112376 can indicate pelvic RT.

The aforementioned CpG site cg21585138, which is mapped to the *CISH* gene, was also identified to be highly related to pelvic RT. It was found to have a role in indicating abdominal RT and associated with adverse health outcomes. This specific methylation pattern can be also caused by pelvic RT.

Another methylation site, cg21745092, was identified as a key feature. This CpG site is located on chr8:68868519 and mapped to the *PREX2* gene. The methylation level of cg21745092 is associated with age, suggesting a potential accumulation of lesions and a risk for disease [40]. *PREX2* can reportedly promote cell proliferation, invasion, and migration in pancreatic cancer [61]. *PREX2* plays an important role in regulating RAC activity and also participates in tumor susceptibility and disease progression [62]. We attributed the methylation status of cg21745092 to a certain environmental burden, such as radiation, that increases the risk for disease progression.

Among the decision rules for identifying pelvic RT, the hypermethylation of cg25531874 was found to be involved in many criteria. The CpG site cg25531874 is located on chr19:39440669 and mapped to the *FBXO17* gene. This gene is related to MHC-mediated antigen processing and presentation and innate immune response. Various associations between expression changes in *FBXO17* and immune diseases have been reported [49,63]. The differential gene expression of *FBXO17* has been found in numerous diseases, including breast cancer and gliomas [64,65]. These findings suggested that *FBXO17* can be a biomarker for the risk of disease on the basis of the causal relationship between radiation and cancer progression. The methylation of cg25531874 can also serve as a potential biomarker for pelvic RT.

4.5. Limitations of This Study

This study also had some limitations. First, the feature selection methods: Boruta and mRMR were adopted to conduct this investigation. It was unknown whether they were optimum to process such methylation profiles. To date, lots of feature selection methods have been proposed. Additional essential methylation features, rules and better classifiers can be obtained with other feature selection methods. Second, several classification rules were extracted from each of the four RT datasets. However, we can only obtain elementary methylation patterns for patients who underwent different types of RT. Further deep investigations were still necessary. Finally, as a bioinformatics study, the new findings (methylation sites and rules) have not been validated by traditional experiments. We hope that related investigators can make further validations based on our findings.

5. Conclusions

In conclusion, our study computationally investigated the relationship between DNA methylation and RT. We first used the Boruta and mRMR methods to filter and rank features from four datasets, namely abdominal RT, brain RT, chest RT, and pelvic RT datasets. These feature lists were then sent into the IFS method, which used classification algorithms, such as DF, to find the best number of features and construct the optimal classifiers. Furthermore, decision rules for the quantitative description of the relationship between the methylation site and RT were developed. Several crucial methylation sites were identified to be highly associated with cancer RT, suggesting that RT has a substantial influence on DNA methylation patterns. We also revealed the specific methylation modifications associated with region-specific cancer RT, implying the different effects of radioactive exposures on specific body parts. These findings not only offer fresh insights into the regulatory role of methylation changes in cancer therapy but also provide a useful analytical approach.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/biology11040607/s1>, Table S1: Feature lists were generated from the four methylation datasets (abdominal RT, brain RT, chest RT, and pelvic-RT) after Boruta feature selection and mRMR analysis; Table S2: Performance of different classifiers on the different numbers of methylation features on the abdominal RT dataset; Table S3: Performance of different classifiers on the different numbers of methylation features on the brain RT dataset; Table S4: Performance of different classifiers on the different numbers of methylation features on the chest RT dataset; Table S5: Performance of different classifiers on the different numbers of methylation features on the pelvic RT dataset; Table S6: Classification rules extracted by each of the best DT classifier from the different methylation datasets.

Author Contributions: Conceptualization, L.L., T.H. and Y.C.; methodology, S.D. and K.F.; validation, L.L. and T.H.; formal analysis, Z.L. and W.G.; data curation, T.H.; writing—original draft preparation, Z.L., W.G. and S.D.; writing—review and editing, T.H.; supervision, Y.C. funding acquisition, T.H. and Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Strategic Priority Research Program of Chinese Academy of Sciences [XDB38050200, XDA26040304], National Key R&D Program of China [2018YFC0910403], the Fund of the Key Laboratory of Tissue Microenvironment and Tumor of Chinese Academy of Sciences [202002].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE169156> (accessed on 6 April 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Jacks, T.; Jaffee, E.; Singer, D. *Cancer Moonshot Blue Ribbon Panel Report 2016*; National Cancer Institute: Tokyo, Japan, 2016.
- Kirsch, D.G.; Diehn, M.; Kesarwala, A.H.; Maity, A.; Morgan, M.A.; Schwarz, J.K.; Bristow, R.; Demaria, S.; Eke, I.; Griffin, R.J. The future of radiobiology. *JNCI J. Natl. Cancer Inst.* **2018**, *110*, 329–340. [[CrossRef](#)] [[PubMed](#)]
- Abshire, D.; Lang, M.K. *The Evolution of Radiation Therapy in Treating Cancer, Seminars in Oncology Nursing*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 151–157.
- Oeffinger, K.C.; Mertens, A.C.; Sklar, C.A.; Kawashima, T.; Hudson, M.M.; Meadows, A.T.; Friedman, D.L.; Marina, N.; Hobbie, W.; Kadan-Lottick, N.S. Chronic health conditions in adult survivors of childhood cancer. *N. Engl. J. Med.* **2006**, *355*, 1572–1582. [[CrossRef](#)] [[PubMed](#)]
- Choi, S.J.; Jung, S.W.; Huh, S.; Chung, Y.-S.; Cho, H.; Kang, H. Alteration of DNA methylation in gastric cancer with chemotherapy. *J. Microbiol. Biotechnol.* **2017**, *27*, 1367–1378. [[CrossRef](#)] [[PubMed](#)]
- Relton, C.L.; Smith, G.D. Epigenetic epidemiology of common complex disease: Prospects for prediction, prevention, and treatment. *PLoS Med.* **2010**, *7*, e1000356. [[CrossRef](#)]
- Robertson, K.D. DNA methylation and human disease. *Nat. Rev. Genet.* **2005**, *6*, 597–610. [[CrossRef](#)] [[PubMed](#)]
- Kalinich, J.F.; Catravas, G.N.; Snyder, S.L. The effect of γ radiation on DNA methylation. *Radiat. Res.* **1989**, *117*, 185–197. [[CrossRef](#)]
- Pogribny, I.; Raiche, J.; Slovack, M.; Kovalchuk, O. Dose-dependence, sex-and tissue-specificity, and persistence of radiation-induced genomic DNA methylation changes. *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1253–1261. [[CrossRef](#)]
- Peng, Q.; Weng, K.; Li, S.; Xu, R.; Wang, Y.; Wu, Y. A perspective of epigenetic regulation in radiotherapy. *Front. Cell Dev. Biol.* **2021**, *9*, 261. [[CrossRef](#)]
- Song, N.; Hsu, C.-W.; Pan, H.; Zheng, Y.; Hou, L.; Sim, J.-a.; Li, Z.; Mulder, H.; Easton, J.; Walker, E. Persistent variations of blood DNA methylation associated with treatment exposures and risk for cardiometabolic outcomes in long-term survivors of childhood cancer in the st. Jude lifetime cohort. *Genome Med.* **2021**, *13*, 53. [[CrossRef](#)]
- Kursa, M.B.; Rudnicki, W.R. Feature selection with the boruta package. *J. Stat. Softw.* **2010**, *36*, 1–13. [[CrossRef](#)]
- Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [[CrossRef](#)] [[PubMed](#)]
- Liu, H.; Setiono, R. Incremental feature selection. *Appl. Intell.* **1998**, *9*, 217–230. [[CrossRef](#)]
- Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; Lawrence Erlbaum Associates Ltd.: Mahwah, NJ, USA, 1995; pp. 1137–1145.
- Zhou, Z.-H.; Feng, J. Deep forest. *Natl. Sci. Rev.* **2018**, *6*, 74–86. [[CrossRef](#)] [[PubMed](#)]
- Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
- Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
- Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674. [[CrossRef](#)]
- Zhang, Y.-H.; Li, H.; Zeng, T.; Chen, L.; Li, Z.; Huang, T.; Cai, Y.-D. Identifying transcriptomic signatures and rules for sars-cov-2 infection. *Front. Cell Dev. Biol.* **2021**, *8*, 1763. [[CrossRef](#)]
- Zhang, Y.-H.; Li, Z.; Zeng, T.; Chen, L.; Li, H.; Huang, T.; Cai, Y.-D. Detecting the multiomics signatures of factor-specific inflammatory effects on airway smooth muscles. *Front. Genet.* **2021**, *11*, 599970. [[CrossRef](#)]
- Zhang, Y.-H.; Zeng, T.; Chen, L.; Huang, T.; Cai, Y.-D. Determining protein–protein functional associations by functional rules based on gene ontology and kegg pathway. *Biochim. Biophys. Acta (BBA) Proteins Proteom.* **2021**, *1869*, 140621. [[CrossRef](#)]

23. Pan, X.; Li, H.; Zeng, T.; Li, Z.; Chen, L.; Huang, T.; Cai, Y.-D. Identification of protein subcellular localization with network and functional embeddings. *Front. Genet.* **2021**, *11*, 626500. [[CrossRef](#)]
24. Yang, Y.; Chen, L. Identification of drug–disease associations by using multiple drug and disease networks. *Curr. Bioinform.* **2022**, *17*, 48–59. [[CrossRef](#)]
25. Zhao, X.; Chen, L.; Lu, J. A similarity-based method for prediction of drug side effects with heterogeneous information. *Math. Biosci.* **2018**, *306*, 136–144. [[CrossRef](#)] [[PubMed](#)]
26. Ahmad, F.; Farooq, A.; Khan, M.U.G.; Shabbir, M.Z.; Rabbani, M.; Hussain, I. Identification of most relevant features for classification of francisella tularensis using machine learning. *Curr. Bioinform.* **2020**, *15*, 1197–1212. [[CrossRef](#)]
27. Baranwal, M.; Magner, A.; Elvati, P.; Saldinger, J.; Violi, A.; Hero, A.O. A deep learning architecture for metabolic pathway prediction. *Bioinformatics* **2019**, *36*, 2547–2553. [[CrossRef](#)]
28. Chen, L.; Li, Z.; Zhang, S.; Zhang, Y.-H.; Huang, T.; Cai, Y.-D. Predicting rna 5-methylcytosine sites by using essential sequence features and distributions. *BioMed Res. Int.* **2022**, *2022*, 4035462. [[CrossRef](#)]
29. Liu, H.; Hu, B.; Chen, L.; Lu, L. Identifying protein subcellular location with embedding features learned from networks. *Curr. Proteom.* **2021**, *18*, 646–660. [[CrossRef](#)]
30. Chen, W.; Chen, L.; Dai, Q. Impt-fdnpl: Identification of membrane protein types with functional domains and a natural language processing approach. *Comput. Math. Methods Med.* **2021**, *2021*, 7681497. [[CrossRef](#)]
31. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
32. Matthews, B. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Biophys. Acta (BBA)-Protein Struct.* **1975**, *405*, 442–451. [[CrossRef](#)]
33. Chen, L.; Wang, S.; Zhang, Y.-H.; Li, J.; Xing, Z.-H.; Yang, J.; Huang, T.; Cai, Y.-D. Identify key sequence features to improve crispr sgRNA efficacy. *IEEE Access* **2017**, *5*, 26582–26590. [[CrossRef](#)]
34. Jia, Y.; Zhao, R.; Chen, L. Similarity-based machine learning model for predicting the metabolic pathways of compounds. *IEEE Access* **2020**, *8*, 130687–130696. [[CrossRef](#)]
35. Liang, H.; Chen, L.; Zhao, X.; Zhang, X. Prediction of drug side effects with a refined negative sample selection strategy. *Comput. Math. Methods Med.* **2020**, *2020*, 1573543. [[CrossRef](#)] [[PubMed](#)]
36. Khor, C.C.; Vannberg, F.O.; Chapman, S.J.; Guo, H.; Wong, S.H.; Walley, A.J.; Vukcevic, D.; Rautanen, A.; Mills, T.C.; Chang, K.-C. Cish and susceptibility to infectious diseases. *N. Engl. J. Med.* **2010**, *362*, 2092–2101. [[CrossRef](#)] [[PubMed](#)]
37. Hunter, M.G.; Jacob, A.; O'Donnell, L.C.; Agler, A.; Druhan, L.J.; Coggeshall, K.M.; Avalos, B.R. Loss of ship and cis recruitment to the granulocyte colony-stimulating factor receptor contribute to hyperproliferative responses in severe congenital neutropenia/acute myelogenous leukemia. *J. Immunol.* **2004**, *173*, 5036–5045. [[CrossRef](#)]
38. Joehanes, R.; Just, A.C.; Marioni, R.E.; Pilling, L.C.; Reynolds, L.M.; Mandaviya, P.R.; Guan, W.; Xu, T.; Elks, C.E.; Aslibekyan, S. Epigenetic signatures of cigarette smoking. *Circ. Cardiovasc. Genet.* **2016**, *9*, 436–447. [[CrossRef](#)]
39. Ackermann, M.A.; King, B.; Lieberman, N.A.; Bobbili, P.J.; Rudloff, M.; Berndsen, C.E.; Wright, N.T.; Hecker, P.A.; Kontrogianni-Konstantopoulos, A. Novel obscurins mediate cardiomyocyte adhesion and size via the pi3k/akt/mtor signaling pathway. *J. Mol. Cell. Cardiol.* **2017**, *111*, 27–39. [[CrossRef](#)]
40. Mulder, R.H.; Neumann, A.; Cecil, C.A.; Walton, E.; Houtepen, L.C.; Simpkin, A.J.; Rijlaarsdam, J.; Heijmans, B.T.; Gaunt, T.R.; Felix, J.F. Epigenome-wide change and variation in DNA methylation in childhood: Trajectories from birth to late adolescence. *Hum. Mol. Genet.* **2021**, *30*, 119–134. [[CrossRef](#)]
41. Liu, C.; Marioni, R.E.; Hedman, Å.K.; Pfeiffer, L.; Tsai, P.-C.; Reynolds, L.M.; Just, A.C.; Duan, Q.; Boer, C.G.; Tanaka, T. A DNA methylation biomarker of alcohol consumption. *Mol. Psychiatry* **2018**, *23*, 422–433. [[CrossRef](#)]
42. Joubert, B.R.; Felix, J.F.; Yousefi, P.; Bakulski, K.M.; Just, A.C.; Breton, C.; Reese, S.E.; Markunas, C.A.; Richmond, R.C.; Xu, C.-J. DNA methylation in newborns and maternal smoking in pregnancy: Genome-wide consortium meta-analysis. *Am. J. Hum. Genet.* **2016**, *98*, 680–696. [[CrossRef](#)]
43. Gracilla, D.E.; Korla, P.K.; Lai, M.T.; Chiang, A.J.; Liou, W.S.; Sheu, J.J.C. Overexpression of wild type or a q311e mutant mb21d2 promotes a pro-oncogenic phenotype in hnscc. *Mol. Oncol.* **2020**, *14*, 3065–3082. [[CrossRef](#)]
44. Xu, C.-Z.; Shi, R.-J.; Chen, D.; Sun, Y.-Y.; Wu, Q.-W.; Wang, T.; Wang, P.-H. Potential biomarkers for paclitaxel sensitivity in hypopharynx cancer cell. *Int. J. Clin. Exp. Pathol.* **2013**, *6*, 2745. [[PubMed](#)]
45. Pandi, G.; Nakka, V.P.; Dharap, A.; Roopra, A.; Vemuganti, R. MicroRNA mir-29c down-regulation leading to de-repression of its target DNA methyltransferase 3a promotes ischemic brain damage. *PLoS ONE* **2013**, *8*, e58039. [[CrossRef](#)] [[PubMed](#)]
46. Bonder, M.J.; Kasela, S.; Kals, M.; Tamm, R.; Lokk, K.; Barragan, I.; Buurman, W.A.; Deelen, P.; Greve, J.-W.; Ivanov, M. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. *BMC Genom.* **2014**, *15*, 860. [[CrossRef](#)] [[PubMed](#)]
47. Islam, S.A.; Goodman, S.J.; MacIsaac, J.L.; Obradović, J.; Barr, R.G.; Boyce, W.T.; Kobor, M.S. Integration of DNA methylation patterns and genetic variation in human pediatric tissues help inform ewas design and interpretation. *Epigenet. Chromatin* **2019**, *12*, 1. [[CrossRef](#)] [[PubMed](#)]
48. Dugué, P.A.; Wilson, R.; Lehne, B.; Jayasekara, H.; Wang, X.; Jung, C.H.; Joo, J.E.; Makalic, E.; Schmidt, D.F.; Baglietto, L. Alcohol consumption is associated with widespread changes in blood DNA methylation: Analysis of cross-sectional and longitudinal data. *Addict. Biol.* **2021**, *26*, e12855. [[CrossRef](#)]

49. Gross, A.M.; Jaeger, P.A.; Kreisberg, J.F.; Licon, K.; Jepsen, K.L.; Khosroheidari, M.; Morsey, B.M.; Swindells, S.; Shen, H.; Ng, C.T. Methylome-wide analysis of chronic hiv infection reveals five-year increase in biological age and epigenetic targeting of hla. *Mol. Cell* **2016**, *62*, 157–168. [[CrossRef](#)]
50. Hannum, G.; Guinney, J.; Zhao, L.; Zhang, L.; Hughes, G.; Sada, S.; Klotzle, B.; Bibikova, M.; Fan, J.-B.; Gao, Y. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **2013**, *49*, 359–367. [[CrossRef](#)]
51. Elias, D.; Vever, H.; Laenkholm, A.; Gjerstorff, M.; Yde, C.; Lykkesfeldt, A.; Ditzel, H. Gene expression profiling identifies fyn as an important molecule in tamoxifen resistance and a predictor of early recurrence in patients treated with endocrine therapy. *Oncogene* **2015**, *34*, 1919–1927. [[CrossRef](#)]
52. Xie, Y.-G.; Yu, Y.; Hou, L.-K.; Wang, X.; Zhang, B.; Cao, X.-C. Fyn promotes breast cancer progression through epithelial-mesenchymal transition. *Oncol. Rep.* **2016**, *36*, 1000–1006. [[CrossRef](#)]
53. Yu, B.; Xu, L.; Chen, L.; Wang, Y.; Jiang, H.; Wang, Y.; Yan, Y.; Luo, S.; Zhai, Z. Fyn is required for arhgef16 to promote proliferation and migration in colon cancer cells. *Cell Death Dis.* **2020**, *11*, 652. [[CrossRef](#)]
54. Nones, K.; Waddell, N.; Song, S.; Patch, A.M.; Miller, D.; Johns, A.; Wu, J.; Kassahn, K.S.; Wood, D.; Bailey, P. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of slit- robo, itga2 and met signaling. *Int. J. Cancer* **2014**, *135*, 1110–1118. [[CrossRef](#)] [[PubMed](#)]
55. Ruiz-Ortiz, E.; Iglesias, E.; Soriano, A.; Buján-Rivas, S.; Español-Rego, M.; Castellanos-Moreira, R.; Tomé, A.; Yagüe, J.; Antón, J.; Hernández-Rodríguez, J. Disease phenotype and outcome depending on the age at disease onset in patients carrying the r92q low-penetrance variant in tnfrsf1a gene. *Front. Immunol.* **2017**, *8*, 299. [[CrossRef](#)] [[PubMed](#)]
56. Yang, B.; Pan, Y.-B.; Ma, Y.-B.; Chu, S.-H. Integrated transcriptome analyses and experimental verifications of mesenchymal-associated tnfrsf1a as a diagnostic and prognostic biomarker in gliomas. *Front. Oncol.* **2020**, *10*, 250. [[CrossRef](#)] [[PubMed](#)]
57. Brzozowska, A.; Powrózek, T.; Homa-Mlak, I.; Mlak, R.; Ciesielka, M.; Gołębiowski, P.; Małecka-Massalska, T. Polymorphism of promoter region of tnfrsf1a gene (−610 t > g) as a novel predictive factor for radiotherapy induced oral mucositis in hnc patients. *Pathol. Oncol. Res.* **2018**, *24*, 135–143. [[CrossRef](#)]
58. Eze, I.C.; Jeong, A.; Schaffner, E.; Rezwan, F.I.; Ghantous, A.; Foraster, M.; Vienneau, D.; Kronenberg, F.; Herceg, Z.; Vineis, P. Genome-wide DNA methylation in peripheral blood and long-term exposure to source-specific transportation noise and air pollution: The sapaldia study. *Environ. Health Perspect.* **2020**, *128*, 067003. [[CrossRef](#)]
59. Templin, T.; Paul, S.; Amundson, S.A.; Young, E.F.; Barker, C.A.; Wolden, S.L.; Smilenov, L.B. Radiation-induced micro-rna expression changes in peripheral blood cells of radiotherapy patients. *Int. J. Radiat. Oncol. Biol. Phys.* **2011**, *80*, 549–557. [[CrossRef](#)]
60. Shin, S.; Cha, H.J.; Lee, E.-M.; Jung, J.H.; Lee, S.-J.; Park, I.-C.; Jin, Y.-W.; An, S. Micrnas are significantly influenced by p53 and radiation in hct116 human colon carcinoma cells. *Int. J. Oncol.* **2009**, *34*, 1645–1652.
61. Yang, J.; Gong, X.; Ouyang, L.; He, W.; Xiao, R.; Tan, L. Prex2 promotes the proliferation, invasion and migration of pancreatic cancer cells by modulating the pi3k signaling pathway. *Oncol. Lett.* **2016**, *12*, 1139–1143. [[CrossRef](#)]
62. Casado-Medrano, V.; Baker, M.J.; Lopez-Haber, C.; Cooke, M.; Wang, S.; Caloca, M.J.; Kazanietz, M.G. The role of rac in tumor susceptibility and disease progression: From biochemistry to the clinic. *Biochem. Soc. Trans.* **2018**, *46*, 1003–1012. [[CrossRef](#)]
63. Liu, Y.; Aryee, M.J.; Padyukov, L.; Fallin, M.D.; Hesselberg, E.; Runarsson, A.; Reinius, L.; Acevedo, N.; Taub, M.; Ronninger, M. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **2013**, *31*, 142. [[CrossRef](#)]
64. Gao, G.; Shi, X.; Yao, Z.; Shen, J.; Shen, L. Identification of lymph node metastasis-related micrnas in breast cancer using bioinformatics analysis. *Medicine* **2020**, *99*, e22105. [[CrossRef](#)] [[PubMed](#)]
65. Zeng, W.-J.; Yang, Y.-L.; Wen, Z.-P.; Chen, P.; Chen, X.-P.; Gong, Z.-C. Identification of gene expression and DNA methylation of serpina5 and timp1 as novel prognostic markers in lower-grade gliomas. *PeerJ* **2020**, *8*, e9262. [[CrossRef](#)] [[PubMed](#)]