Article

Nonlinear Techniques and Ridge Regression as a Combined Approach: Carcinoma Identification Case Study

Gerardo Alfonso Perez *[ID] and Raquel Castillo

Biocomp Group, Institute of Advanced Materials (INAM), Universitat Jaume I, 12071 Castelló de la Plana, Spain
* Correspondence: ga284@cantab.net

Abstract: As more genetic information becomes available, such as DNA methylation levels, it becomes increasingly important to have techniques to analyze such data in the context of cancers such as anal and cervical carcinomas. In this paper, we present an algorithm that differentiates between healthy control patients and individuals with anal and cervical carcinoma, using as an input DNA methylation data. The algorithm used a combination of ridge regression and neural networks for the classification task, achieving high accuracy, sensitivity and specificity. The relationship between methylation levels and carcinoma could in principle be rather complex, particularly given that a large number of CpGs could be involved. Therefore, nonlinear techniques (machine learning) were used. Machine learning techniques (nonlinear) can be used to model linear processes, but the opposite (linear techniques simulating nonlinear models) would not likely generate accurate forecasts. The feature selection process is carried out using a combination of prefiltering, ridge regression and nonlinear modeling (artificial neural networks). The model selected 13 CpGs from a total of 450,000 CpGs available per patient with 171 patients in total. The model was also tested for robustness and compared to other more complex models that generated less precise classifications. The model obtained (testing dataset) an accuracy, sensitivity and specificity of 97.69%, 95.02% and 98.26%, respectively. The reduction of the dimensionality of the data, from 450,000 to 13 CpGs per patient, likely also reduced the likelihood of overfitting, which is a very substantial risk in this type of modelling. All 13 CpGs individually generated classification forecasts less accurate than the proposed model.

Keywords: anal cancer; cervical cancer; algorithm

MSC: 65F30

1. Introduction and Literature Review

Some recent articles, such as Deshmukh et al. [1], have estimated that the incidence of anal carcinoma is increasing at 2.7% per year. They also estimated a similar trend for mortality. Similar results were found by Eng et al. [2]. They estimated a 3.1% increase in the mortality rate. Anal and cervical carcinomas are not yet well understood [3–5]. Articles, such as Melbye and Sprogel [6] and Rabkin et al. [7], have mentioned that anal and cervical cancers have common risk factors and other similarities. Parallels between these two illnesses have been mentioned in the existing literature for decades [8–10]. There is increasing research pointing to a link between anal and cervical carcinomas and the human papillomavirus (HPV) with causal relationship or a strong link mentioned in several articles, such as Darrangh and Winkler [11], Franceschi and De Vuyst [12], Škamperle et al. [13] and Ryan et al. [14]. De Sanjose et al. [15] mentions that HPV has been established as a "central and necessary cause of cervical cancer". Immunosuppressed patients, such as HIV patients, have a higher likelihood of developing this type of cancer [16]. Cancer is in fact a common comorbidity in HIV patients [17–19].

Varnani et al. [20] found a sensitivity and specificity of 93.6% and 80.0%, respectively, in a histological analysis of biopsies of suspected anal carcinoma patients. Van der Zee et al. [21] found a similar specificity (79%) when modeling the risk of anal carcinoma in

HIV-positive patients using as an input DNA methylation data. Other authors have found similar results [22,23].

Alterations of DNA methylation in anal carcinoma have been mentioned in several articles, such as Zhang et al. [24]. The authors of this paper concluded that aberrant methylation is frequent in anal carcinomas. Some articles, such as Siegel et al. [25], have studied changes in methylation levels in both cervical and anal carcinomas, finding also changes in the methylation patterns. Machine learning techniques [26] are an increasingly important tool in many non-medical [27,28] and medical research areas [29–31], and cancer research in no exception [32–34]. Some authors, such as Cuocolo et al. [35], have mentioned that machine learning "could become an essential part of…oncological screening". Other authors such as Forsch et al. [36] and Kourou et al. [37] have concluded similarly. There are some interesting articles applying machine learning techniques in the context of carcinomas. For instance, Huang et al. [38] used deep neural networks applied to DNA methylation data aiming to predict outcomes for patients. Nartowt et al. [39] applied an artificial neural network approach for scoring colorectal cancer using self-reported personal health data, achieving a sensitivity and specificity of 57% and 89%, respectively. Methylation data have been used in the analysis of other cancers, such as lung carcinomas (Marchevsky [40], Ligor et al. [41]), glioblastoma (Calabrese et al. [42]), endometrial cancer (Pergialiotis et al. [43]) and gastric cancer (Zhang et al. [44]). Lin et al. [45] used a LASSO approach, which is a special case of ridge regression, in the analysis of the relationship between the expression of m6A RNA methylation and hepatocellular carcinoma prognosis. Butcher and Beck [46] also used a LASSO approach in the context of colon cancer (but no machine learning techniques such as neural networks). Zhong et al. [47] also used the LASSO approach and concluded that this approach with linear regression models has limited prediction power. Cancer screening methods for anal carcinoma (e.g., occult blood test) and cervical carcinoma (e.g., pap smear) are well established. Methylation changes might be able to be detected (but this would need to be tested by further experimental data) before there is occult blood. It can also potentially be used for targeted medicine, i.e., DNA methylation profiles can potentially be used to try to assign more suitable treatment options, according to their methylation profile, to patients.

There are several articles in the existing literature highlighting the applicability of artificial neural networks in the context of nonlinear processes. For example, Zhang et al. [48] applied this technique to nonlinear time series. Liu et al. [49] proposed a multilevel artificial neural network nonlinear equalizer for millimeter-wave mobile fronthaul systems. There are also several papers related to nonlinear control processes, see for instance Cong et al. [50].

There are other ways to carry out this type of analysis. For instance, it is possible to use logistic regression [51] instead of artificial neural networks. There are advantages and disadvantages of using these techniques. Tu [52] mentioned that one of the main advantages of artificial neural networks is their ability to implicitly detect complex nonlinear relationships as well as the ability to detect all possible combinations between predictor variables. One of the disadvantages mentioned by Tu when comparing artificial neural networks and logistic regression was the black box behavior of artificial neural networks with some of the models created being potentially very complex and difficult to interpret.

Objectives

The main objective of this paper is to distinguish between healthy control patients and patients with anal or cervical carcinoma using DNA methylation data and an algorithm combining ridge regression with nonlinear techniques, such as artificial neural networks.

2. Materials and Methods

2.1. Data

The data were obtained from the GEO database with accession code GSE 186859 (publicly available), containing 171 samples of genomic DNA, of which 152 are anal and cervical carcinomas as well as pre-tumours (AIN3 with 13 cases and CIN3 with 9 cases),

and the rest are control healthy patients. The dataset consists of 28 cervical samples and 143 anal samples. The data were obtained using the standard illumina protocol, and the chips were scanned on a HiScanSQ System. The researchers that collected the data preprocessed it by performing background correction and normalization using the minfi Bioconductor software in R. Given the relatively low number of pre-tumor cases, the tumor and pre-tumor cases were combined into a single category, which assumes that pre-tumors and tumors have altered DNA methylation levels compared to a healthy individual. There are approximately 450,000 CpGs per patient.

2.2. Notation

The CpG methylation data (X) are represented (Equation (1)) in a matrix form [53]:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \cdots & x_{mn} \end{pmatrix} \tag{1}$$

Each column represents the methylation data for a given patient, with each row representing the same CpG across different patients. The methylation level is a percentage value expressed as a number ranging from 0 (not methylated) to 1 (fully methylated). As an example of this nomenclature, X_{21} represents the methylation data for patient 1 in CpG 2. It is also convenient to have a vector (Equation (2)) distinguishing between control and patients ($Y_i = \{0, 1\}$).

$$Y = \{y_1, y_2, \ldots, y_n\} \tag{2}$$

2.3. Preliminary Filtering

As usual in nonlinear models, the data need to be divided into a training and a testing dataset. The testing dataset contains approximately 20% of the total data. Furthermore, 10% of the data (training dataset) were used as validation data. We carried out cross-validation 10 times. There are several interesting papers covering validation, see for instance [54,55]. The testing dataset was not used during the training phase. The reported measures, such as accuracy, sensitivity and specificity, are those obtained in the testing dataset (unused duting the trainign phase). A preliminary step consists in filtering each CpG ($X^t = \{x_{t1}, x_{t2}, x_{t3}, \ldots, x_{tn}\}$) individually using binomial regression. In this regression, the independent variable is the methylation level for each CpG across all patients in the training dataset, and the dependent variable is Y (Equation (2)). In this first step, all the CpGs with a p-value bigger than 0.05 were excluded from the analysis.

2.4. Variance Filtering

After this preliminary filtering, an additional filtering was carried out. In this step, the k CpGs with the highest variance were selected. The idea behind this approach is that in the extreme, a CpG that does not have any variation would not be useful as an input for an algorithm that tries to distinguish between control cases and patients.

2.5. Combined Ridge Regression and Nonlinear Modeling

It is possible to further reduce the dimensionality of the data using an approach such as ridge regression [56–58]. This approach automatically reduces the dimensionality of the data by making some of the coefficients in the regression equal to zero. The number of coefficients made equal to zero depends on the parameter α in the ridge regression. In principle, there is no indication that the relationship between the level of DNA methylation and the presence or absence of a tumor should follow a linear relationship. Hence, a nonlinear approach (artificial neural networks) was followed. In this way, the ridge regression selects

the CpGs that are then used as inputs in the nonlinear model. The neural network accuracy will depend on factors such as the number of neurons (l) used. Hence, we have the following optimization problem (Equations (3)–(5)). The artificial neural network uses a scaled conjugate gradient backpropagation as a training algorithm, a hidden layer consisting of a hyperbolic tangent sigmoid and an output layer with a softmax transfer function. The training algorithm was used only with the training dataset.

$$\max_{l^*, \alpha^*} \quad f(l, \alpha) \tag{3}$$

$$\text{s.t.} \quad l \leq l_{max}, \tag{4}$$

$$0 < \alpha \leq 1. \tag{5}$$

where l is the number of neurons in the artificial neural network, α is the $\alpha - parameter$ in the ridge regression, and f is a function measuring the goodness of the binary forecast (patient vs. control) of the model output compared to the actual values. This function (f) can be for example the accuracy of the model or the sensitivity or specificity of the model. This task can be performed following a grid approach (Algorithm 1):

Algorithm 1 Grid approach optimization (l_i, α_j)

Input: l_i, α_j
Output: $f_{ij}(l_i, \alpha_j)$
1. Create a grid of values for $l_i = \{l_1, l_2, l_3, \ldots, l_{max}\}$
2. Create a grid of values for $\alpha_j = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_{max}\}$
3. Estimate forecast (F) of the status of patients $F = F_{ij}(l_i, \alpha_j)$
4. Estimate goodness of fit to the binary classification $f = f_{ij}(l_i, \alpha_j)$
5. Repeat steps 3 and 4 q times and obtain mean values
6. Select

$$\sup \quad \left(\frac{1}{q} \sum_{s=1}^{q} f_{ij}^s - g(i) \right) = \bar{f}_{ij}^*$$

$$\text{s.t.} \quad l \leq l_{max},$$
$$0 < \alpha \leq 1.$$

where $g(i)$ is a penalty function of the type $g(i) = \beta \cdot i$

With the type of approach presented, it is also necessary to carry out a robustness analysis in which, after $\bar{f}_{ij}^*$ is obtained (and hence i and j fixed), the modeling needs to be repeated r times. This step is necessary given the random initialization of the weights in neural networks that result in different outputs, even if the inputs and the structure of the neural network remain unchanged. The value k (variance filtering) needs to be chosen in order for the grid approach to be computationally feasible. Another important step is modelling each CpG individually ($x^t = \{x_{t1}, x_{t2}, x_{t3}, \ldots, x_{tn}, \}$) to study the potential case in which any of the CpGs individually might generate results comparable to the previously generated model.

An alternative to Algorithm 1, in which the optimization is carried out on the number of neurons (l_i) and the $\alpha - factor$ (α_i) of the ridge regression, would be to expand it to include a variable number of layers (κ) as well as adding different types of penalty functions. This can be seen in Algorithm 2.

The purpose of the penalty function is to penalize overly complex model structures that could potentially reduce the generalization capability of the model.

Algorithm 2 Grid approach optimization (l_i, α_j, κ_u)

Input: l_i, α_j, κ_u
Output: $f_{iju}(l_i, \alpha_j, \kappa_u)$
1. Create a grid of values for $l_i = \{l_1, l_2, l_3, \ldots, l_{max}\}$
2. Create a grid of values for $\alpha_j = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_{max}\}$
3. Create a grid of values for $\kappa_u = \{\kappa_1, \kappa_2, \kappa_3, \ldots, \kappa_{max}\}$
4. Estimate forecast (F) of the status of patients $F = F_{iju}(l_i, \alpha_j, \kappa_u)$
5. Estimate goodness of fit to the binary classification $f = f_{iju}(l_i, \alpha_j, \kappa_u)$
6. Repeat steps 4 and 5 q times and obtain mean values
7. Select

$$\sup \quad \left(\frac{1}{q} \sum_{s=1}^{q} f_{iju}^s - g(i) \right) = \bar{f}_{iju}^*$$

$$\text{s.t.} \quad l \leq l_{max},$$
$$0 < \alpha \leq 1,$$
$$\kappa \leq \kappa_{max}.$$

where in this case, the penalty function can be $g(i, \kappa) = \beta_1 \cdot i + \beta_2 \cdot \kappa$ or a quadratic expression $g(i, \kappa) = \beta_1 \cdot i^2 + \beta_2 \cdot \kappa^2$

---

3. Results

After the initial pre-filtering (excluding CpGs with a p value bigger than 0.05), the 200 CpGs ($\kappa = 200$) with the highest variance were selected. As previously mentioned, the assumption is that CpGs with no or very little variance will be of limited use as an input for a classification algorithm. The value of $\kappa = 200$ was selected in order to make the calculations computationally feasible while at the same time maintaining a relatively high number of CpGs. Then, Algorithm 1 was applied to the filtered data (containing 200 CpGs per patient). As described in Section 2, the algorithm tries to find a suitable combination of number of CpGs, which are a function of the α parameter in the ridge regression, and the number of neurons. For clarity purposes, in Figure 1, a graph can be seen showing the results for a given number of neurons and the accuracy at the different α values. A sample of the goodness of the model for a specific configuration can be seen in the ROC curves in Figure 2.

Algorithm 1 then expands this approach for a grid of different numbers of neurons, as can be seen in Figure 3. This approach resulted in a model with only 13 CpGs selecting an accuracy of 97.69%. The specificity and sensitivity of the model were 98.26% and 95.02%, respectively. The number of neurons (l) selected was 790. The average methylation level for these 13 CpGs (for control and patients) can be seen in Figure 4. The list of these 13 CpGs can be found in the Appendix A (Table A1).

It is important to obtain a robust model in which the results are hopefully repeatable. In order to test the robustness of the model, the simulation was repeated 1000 times with the same inputs and network structure. The random initialization of the weights leads to changes in the classification forecast of the model even with the same inputs and network structure. In Figure 5, a histogram can be found showing the resulting accuracy of these simulations. It can seem that it is relatively tightly centered with no frequent outliers. It is also important to analyze each of these CpGs individually. No single CpG has a mean accuracy above 88.94%. Accuracy for each CpG (individually) can be seen in Figure 6.
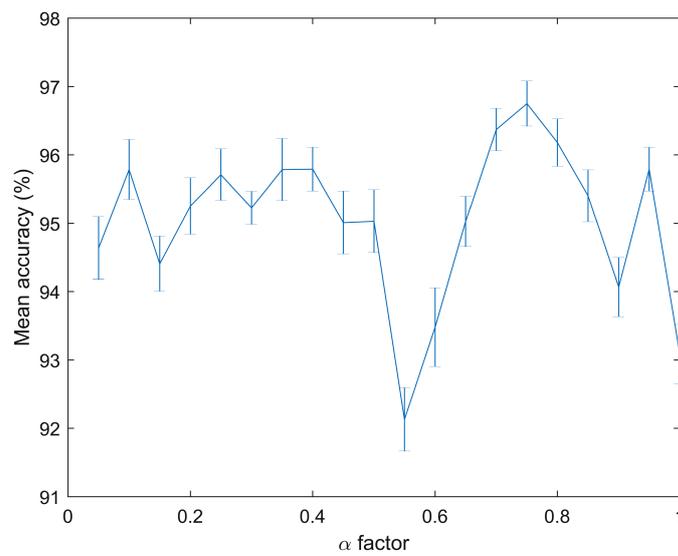
Figure 1. Graph showing the accuracy obtained when fixing the number of neurons and changing the α factor.
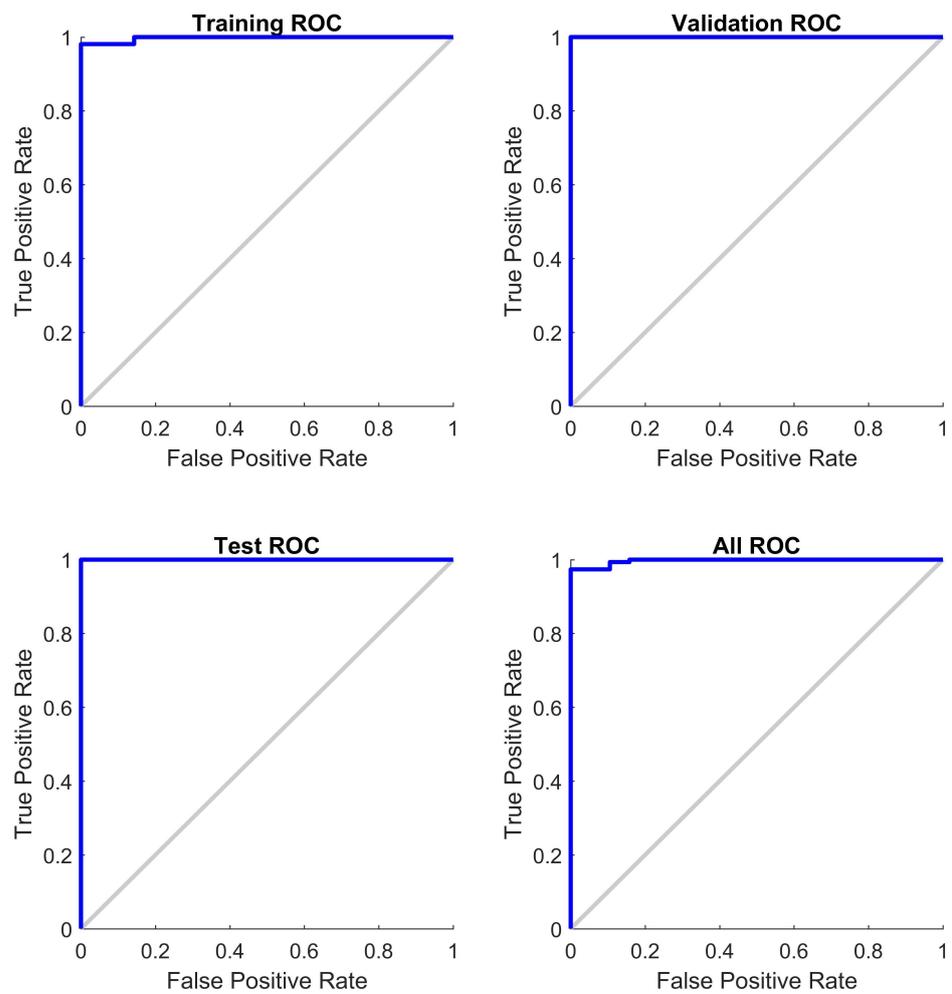

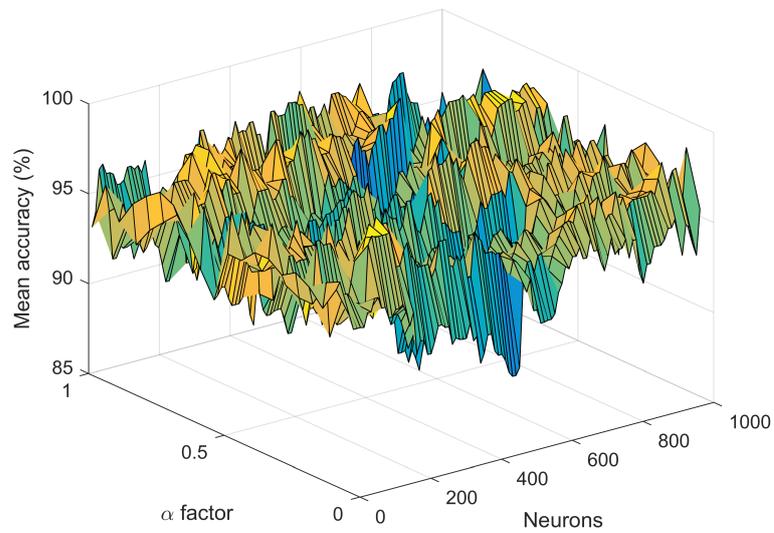
Figure 2. ROC sample curve for one of the estimations.

Figure 3. Accuracy obtained using Algorithm 1 (grid approach varying the number of neurons and α factor in a grid).
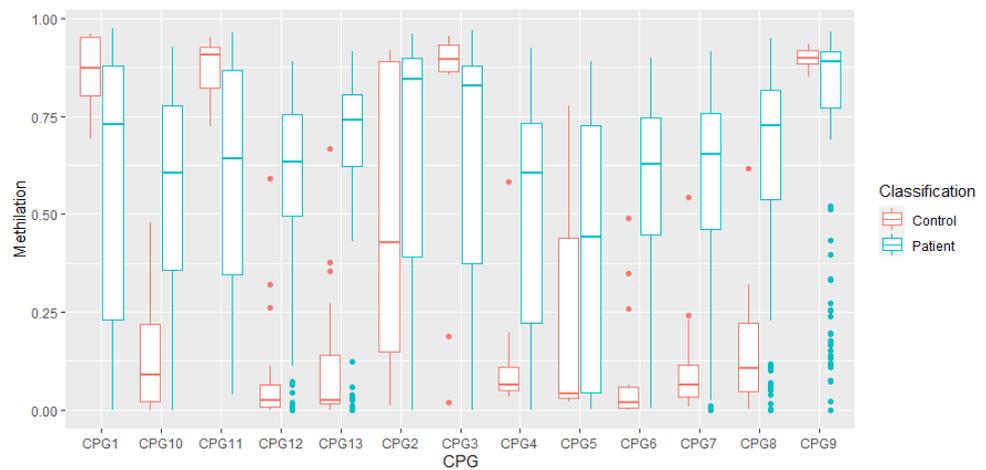

Figure 4. Mean methylation values for patients and control cases.
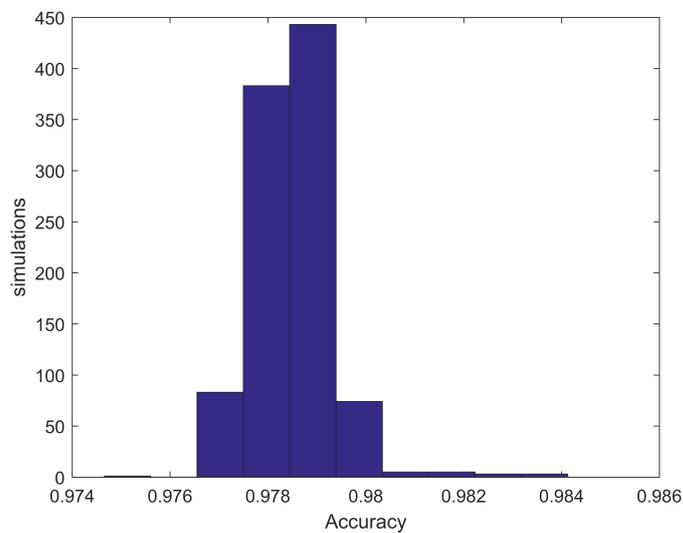

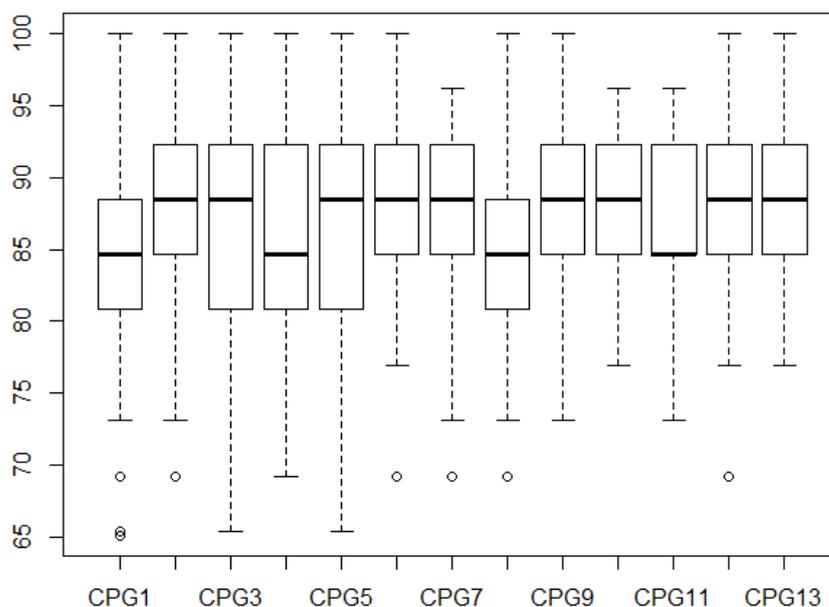**Figure 5.** Histogram of the accuracy obtained in 1000 simulations.

Figure 6. Classification accuracy (%) of each CpG individually.

In Table 1, the results for Algorithms 1 and 2 can be seen. One of the main differences between Algorithms 1 and 2 is that in Algorithm 2, the number of layers was also modified, and two penalty functions were used. The results of the second algorithm were slightly less precise than those in the first algorithm. The best results using Algorithm 1 were with one hidden layer, 790 neurons (with the penalty function $g(i, \kappa) = \beta_1 \cdot i + \beta_2 \cdot \kappa$), and with two hidden layers, 840 neurons (with the quadratic penalty function $g(i, \kappa) = \beta_1 \cdot i^2 + \beta_2 \cdot \kappa^2$). The base case, using all the CpGs and no optimization, is also shown for comparison purposes. The excessive number of inputs in this base case (no filtering) might cause overfitting in the model.

Table 1. Metrics comparing the results of the algorithms.

Metric	Algorithm 1	Algorithm 2 *	Algorithm 2 **	Base
Accuracy	97.69	96.92	94.62	69.23
Specificity	98.26	97.34	98.26	78.95
Sensitivity	95.02	93.33	78.67	42.86

* Algorithm 2 with linear penalty function. ** Algorithm 2 with quadratic penalty function.

4. Discussion

The proposed approach of using DNA methylation data, as inputs, and an algorithm combining ridge regression and artificial neural networks, for the task of differentiating between healthy control individuals and individuals with anal and cervical carcinomas, generated accurate results with specificity and sensitivity higher than ones obtained in other papers in the field. The algorithm selected 13 CpGs from a starting point of approximately 450,000 CpGs per patient. Technological developments have made it possible to obtain such large amounts of methylation data but at the same time have made the analysis of such data challenging. Given that there is no indication that there is a linear relationship between the level of methylation (CpGs) and the presence of anal or cervical carcinoma, the modeling approach was performed with nonlinear techniques such as artificial neural networks. One of the issues with this type of model is the risk of overfitting, particularly in this type of situation in which there is a large number of inputs per patient but a smaller number of patients. In order to reduce this type of risk, it is important to reduce the dimensionality of the data. Additionally, this reduction in the dimensionality can point to CpGs that might be important as biomarkers in the context of the disease. The selected model was tested for robustness, with the classification estimates remaining accurate for the vast majority of

the simulations. No individual CpGs, of those 13 selected by the model, achieved a mean accuracy above 88.94%, which is substantially lower than the 97.69% accuracy obtained by the model. Increasing the complexity of the models, by for instance adding more layers to the neural network, did not appear to increase the accuracy of the model. This might be again related to the issue of overfitting. Similarly, adding more complex penalty functions, such as for instance a quadratic function rather than a linear function, did not improve the accuracy.

Limitations and Future Work

There are some limitations in this analysis. For instance, there were only 171 patients analyzed. While the number of patients is not too small, this type of analysis would benefit from a larger cohort of patients. As more data become available, this type of approach can be retested with larger cohorts. Given the larger number of cases of anal carcinoma compared to cervical carcinoma, it is likely that the model will be more precise when classifying anal carcinomas. While there is a clear protocol for obtaining DNA methylation data, there are will always be some small differences in the way that different laboratories collect and present the data. These experimental differences could result in differences in the DNA methylation data and hence reduce the accuracy (and other metrics). It would be very interesting to have time evolution data for the patients that have carcinomas as well as their treatments. It is conceivable that treatment of the patients could potentially be individualized according to their methylation profile, but there is currently, to the best of our knowledge, no available data to actually test this hypothesis. This could be a very interesting area of future research with direct clinical applications.

5. Conclusions

The proposed approach is able to generate an accuracy, sensitivity and specify of classification forecasts of 97.69%, 95.02% and 98.26%, respectively, illustrating that a combination of DNA methylation with nonlinear methods such as artificial neural networks might be useful in the task of identifying patients with a carcinoma. This approach could be complementary to the existing techniques such as occult blood test and pap smear. This is conceivable, but additional testing would be required to support this hypothesis, that DNA methylation changes might be present in the patient before there are clinical indications (occult blood test). This is an important research question that should be addressed in future research. Additionally, it is possible that finding different DNA methylation signatures could be used for personalized treatments. This is another area in which more research would be needed. The model achieved a substantial reduction in the number of CpGs used as input from a starting point of approximately 450,000 to only 13. This is important, as having an excessively large number of inputs could lead to overfitting issues. The combination of these 13 CpGs generated more accurate forecasts that any of them individually. The list of these 13 CpGs can be found in the Appendix A.

Appendix A

List of 13 CpGs selected by Algorithm 1.

Table A1. CpG obtained using Algorithm 1.

CpG	CpG Code (GEO)
1	cg15290312
2	cg14331362
3	cg01270299
4	cg07352438
5	cg19393008
6	cg26110710
7	cg21523564
8	cg14487131
9	cg00259849
10	cg14262681
11	cg02263377
12	cg06073449
13	cg18456523

References

1. Deshmukh, A.A.; Suk, R.; Shiels, M.S.; Sonawane, K.; Nyitray, A.G.; Liu, Y.; Gaisa, M.M.; Palefsky, J.M.; Sigel, K. Recent trends in squamous cell carcinoma of the anus incidence and mortality in the United States, 2001–2015. *JNCI J. Natl. Cancer Inst.* **1991**, *338*, 657–659. [CrossRef] [PubMed]
2. Eng, C.; Ciombor, K.K.; Cho, M.; Dorth, J.A.; Rajdev, L.N.; Horowitz, D.P.; Gollub, M.J.; Jacome, A.A.; Lockney, N.A.; Muldoon, R.L. Anal cancer: Emerging standards in a rare disease. *J. Clin. Oncol.* **2022**, *40*, 2774–2788. [CrossRef] [PubMed]
3. Monsrud, A.L.; Avadhani, V.; Mosunjac, M.B.; Flowers, L.; Krishnamurti, U. Programmed death ligand-1 expression is associated with poorer survival in anal squamous cell carcinoma. *Arch. Pathol. Lab. Med.* **2022**, *146*, 1094–1101. [CrossRef] [PubMed]
4. Saiki, Y.; Yamada, K.; Tanaka, M.; Fukunaga, M.; Irei, Y.; Suzuki, T. Prognosis of anal canal adenocarcinoma versus lower rectal adenocarcinoma in Japan: A propensity score matching study. *Surg. Today* **2022**, *52*, 420–430. [CrossRef] [PubMed]
5. Lupi, M.; Brogden, D.; Howell, A.; Tekkis, P.; Mills, S.; Kontovounisios, C. Anal Cancer in High-Risk Women: The Lost Tribe. *Cancers* **2022**, *15*, 60. [CrossRef]
6. Melbye, M.; Sprogel, P. Aetiological parallel between anal cancer and cervical cancer. *Lancet* **1991**, *338*, 657–659. [CrossRef]
7. Rabkin, C.S.; Biggar, R.J.; Melbye, M.; Curtis, R.E. Second primary cancers following anal and cervical carcinoma: Evidence of shared etiologic factors. *Am. J. Epidemiol.* **1992**, *136*, 54–58. [CrossRef]
8. Scholefield, J.H.; Talbot, I.C.; Whatrup, C.; Sonnex, C.; Palmer, J.G.; Mindel, A.; Northover, J.M.A. Anal and cervical intraepithelial neoplasia: Possible parallel. *Lancet* **1989**, *334*, 765–769. [CrossRef]
9. Palmer, J.G.; Scholffield, J.H.; Coates, P.J.; Shepherd, N.A.; Jass, J.R.; Crawford, L.V.; Northover, J.M.A. Anal cancer and human papillomaviruses. *Dis. Colon Rectum* **1989**, *32*, 1016–1022. [CrossRef]
10. Doggett, S.W.; Green, J.P.; Cantril, S.T. Efficacy of radiation therapy alone for limited squamous cell carcinoma of the anal canal. *Int. J. Radiat. Oncol. Biol. Phys.* **1988**, *15*, 1069–1072. [CrossRef]
11. Darragh, T.M.; Winkler, B. Anal cancer and cervical cancer screening: Key differences. *Cancer Cytopathol.* **2011**, *119*, 5–19. [CrossRef] [PubMed]
12. Franceschi, S.; De Vuyst, H. Human papillomavirus vaccines and anal carcinoma. *Curr. Opin. HIV AIDS* **2009**, *4*, 57–63. [CrossRef] [PubMed]
13. Škamperle, M.; Kocjan, B.J.; Maver, P.J.; Seme, K.; Poljak, M. Human papillomavirus (HPV) prevalence and HPV type distribution in cervical, vulvar, and anal cancers in central and eastern Europe. *Acta Dermatovenerol. Alpina Panon. Adriat.* **2013**, *22*, 1–5. [PubMed]
14. Ryan, D.P.; Compton, C.C.; Mayer, R.J. Carcinoma of the anal canal. *N. Engl. J. Med.* **2000**, *342*, 792–800. [CrossRef]
15. de Sanjose, S.; Bruni, L.; Alemany, L. HPV in genital cancers (at the exception of cervical cancer) and anal cancers. *La Presse Médicale* **2014**, *43*, 423–428. [CrossRef]
16. Williams, G.R.; Talbot, I.C. Anal carcinoma—A histological review. *Histopathology* **1994**, *25*, 507–516. [CrossRef]
17. Sumner, L.; Kamitani, E.; Chase, S.; Wang, Y. A systematic review and meta-analysis of mortality in anal cancer patients by HIV status. *Histopathology* **2022**, *76*, 102069. [CrossRef]
18. Naito, T.; Suzuki, M.; Fukushima, S.; Yuda, M.; Fukui, N.; Tsukamoto, S.; Fujibayashi, K.; Goto-Hirano, K.; Kuwatsuru, R. Comorbidities and co-medications among 28 089 people living with HIV: A nationwide cohort study from 2009 to 2019 in Japan. *HIV Med.* **2022**, *23*, 485–493. [CrossRef]

19. Muchengeti, M.; Bartels, L.; Olago, V.; Dhokotera, T.; Chen, W.C.; Spoerri, A.; Rohner, E.; Butikofer, L.; Ruffieux, Y.; Singh, E. Cohort profile: The South African HIV Cancer Match (SAM) Study, a national population-based cohort. *BMJ Open* **2022**, *12*, 053460. [CrossRef]

20. Varnai, A.D.; Bollmann, M.; Griefingholt, H.; Speich, N.; Schmitt, C.; Bollmann, R.; Decker, D. HPV in anal squamous cell carcinoma and anal intraepithelial neoplasia (AIN) Impact of HPV analysis of anal lesions on diagnosis and prognosis. *Int. J. Color. Dis.* **2006**, *21*, 135–142. [CrossRef]

21. van der Zee, R.P.; Richel, O.; van Noesel, C.J.M.; Novianti, P.W.; Ciocanea-Teodorescu, I.; van Splunter, A.P.; Duin, S.; van den Berk, G.E.L.; Meijer, C.; Quint, W. Host cell deoxyribonucleic acid methylation markers for the detection of high-grade anal intraepithelial neoplasia and anal cancer. *Clin. Infect. Dis.* **2019**, *68*, 1110–1117. [CrossRef] [PubMed]

22. Legarth, R.; Helleberg, M.; Kronborg, G.; Larsen, C.S.; Pedersen, G.; Pedersen, C.; Jensen, J.; Nielsen, L.N.; Gerstoft, J.; Obel, N. Anal carcinoma in HIV-infected patients in the period 1995–2009: A Danish nationwide cohort study. *Scand. J. Infect. Dis.* **2013**, *45*, 453–459. [CrossRef] [PubMed]

23. Kreuter, A.; Potthoff, A.; Brockmeyer, N.H.; Gambichler, T.; Swoboda, J.; Stucker, M.; Schmitt, M.; Pfister, H.; Wieland, U. Anal carcinoma in human immunodeficiency virus-positive men: Results of a prospective study from Germany. *Br. J. Dermatol.* **2010**, *162*, 1269–1277. [CrossRef] [PubMed]

24. Zhang, J.; Martins, C.R.; Fansler, Z.B.; Roemer, K.L.; Kincaid, E.A.; Gustafson, K.S.; Heitjan, D.F.; Clark, D.P. DNA methylation in anal intraepithelial lesions and anal squamous cell carcinoma. *Clin. Cancer Res.* **2005**, *11*, 6544–6549. [CrossRef]

25. Siegel, E.M.; Ajidahun, A.; Berglund, A.; Guerrero, W.; Eschrich, S.; Putney, R.M.; Magliocco, A.; Riggs, B.; Winter, K.; Simko, J.P. Genome-wide host methylation profiling of anal and cervical carcinoma. *PLoS ONE* **2021**, *16*, e0260857. [CrossRef]

26. Greener, J.G.; Kandathil, S.M.; Moffat, L.; Jones, D.T. A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 40–55. [CrossRef]

27. Salau, A.O.; Jain, S. Feature extraction: A survey of the types, techniques, applications. In Proceedings of the 2019 International Conference on Signal Processing and Communication (ICSC), Noida, India, 7–9 March 2019; pp. 158–164.

28. Guarino, A.; Lettieri, N.; Malandrino, D.; Zaccagnino, R.; Capo, C. Adam or Eve? Automatic users' gender classification via gestures analysis on touch devices. *Neural Comput. Appl.* **2022**, *34*, 18473–18495. [CrossRef]

29. Rabbani, N.; Kim, G.Y.; Suarez, C.J.; Chen, J.H. Applications of machine learning in routine laboratory medicine: Current state and future directions. *Clin. Biochem.* **2021**, *103*, 1–7. [CrossRef]

30. Quazi, S. Artificial intelligence and machine learning in precision and genomic medicine. *Med. Oncol.* **2022**, *39*, 120. [CrossRef]

31. Mueller, B.; Kinoshita, T.; Peebles, A.; Graber, M.A.; Lee, S. Artificial intelligence and machine learning in emergency medicine: A narrative review. *Acute Med. Surg.* **2022**, *9*, 740. [CrossRef]

32. Cai, Z.; Poulos, R.C.; Liu, J.; Zhong, Q. GMachine learning for multi-omics data integration in cancer. *iScience* **2022**, *2022*, 103798. [CrossRef] [PubMed]

33. Capobianco, E. High-dimensional role of AI and machine learning in cancer research. *Br. J. Cancer* **2022**, *126*, 523–532. [CrossRef] [PubMed]

34. Painuli, D.; Bhardwaj, S. Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review. *Comput. Biol. Med.* **2022**, *2022*, 105580. [CrossRef]

35. Cuocolo, R.; Caruso, M.; Perillo, T.; Ugga, L.; Petretta, M. Machine learning in oncology: A clinical appraisal. *Cancer Lett.* **2020**, *481*, 55–62. [CrossRef]

36. Forsch, S.; Klauschen, F.; Hufnagl, P.; Roth, W. Artificial intelligence in pathology. *Deutsches Ärzteblatt Int.* **2021**, *118*, 199. [CrossRef]

37. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17. [CrossRef]

38. Huang, G.; Wang, C.; Fu, X. Bidirectional deep neural networks to integrate RNA and DNA data for predicting outcome for patients with hepatocellular carcinoma. *Future Oncol.* **2021**, *17*, 4481–4495. [CrossRef] [PubMed]

39. Nartowt, B.J.; Hart, G.R.; Roffman, D.A.; Llor, X.; Ali, I.; Muhammad, W.; Liang, Y.; Deng, J. Scoring colorectal cancer risk with an artificial neural network based on self-reportable personal health data. *PLoS ONE* **2019**, *14*, 0221421. [CrossRef]

40. Marchevsky, A.M. The Use of Artificial Neural Networks for the Diagnosis and Estimation of Prognosis in Cancer Patients. *Outcome Predict. Cancer* **2007**, 243–259. [CrossRef]

41. Ligor, T.; Pater, L.; Buszewski, B. Application of an artificial neural network model for selection of potential lung cancer biomarkers. *J. Breath Res.* **2015**, *9*, 027106. [CrossRef]

42. Calabrese, E.; Rudie, J.D.; Rauschecker, A.M.; Villanueva-Meyer, J.E.; Clarke, J.L.; Solomon, D.A.; Cha, S. Combining radiomics and deep convolutional neural network features from preoperative MRI for predicting clinically relevant genetic biomarkers in glioblastoma. *Neuro-Oncol. Adv.* **2022**, *4*, 60. [CrossRef] [PubMed]

43. Pergialiotis, V.; Pouliakis, A.; Parthenis, C.; Damaskou, V.; Chrelias, C.; Papantoniou, N.; Panayiotides, I. The utility of artificial neural networks and classification and regression trees for the prediction of endometrial cancer in postmenopausal women. *Public Health* **2018**, *164*, 1–6. [CrossRef] [PubMed]

44. Zhang, G.; Xue, Z.; Yan, C.; Wang, J.; Luo, H. A novel biomarker identification approach for gastric cancer using gene expression and DNA methylation dataset. *Front. Genet.* **2021**, *12*, 644378. [CrossRef] [PubMed]

45. Lin, Y.; Yao, Y.; Wang, Y.; Wang, L.; Cui, H. PD-L1 and immune infiltration of m6A RNA methylation regulators and its miRNA regulators in hepatocellular carcinoma. *BioMed Res. Int.* **2021**, *2021*, 1–16. [CrossRef]
46. Butcher, L.M.; Beck, S. Probe Lasso: A novel method to rope in differentially methylated regions with 450 K DNA methylation data. *Methods* **2015**, *72*, 21–28. [CrossRef] [PubMed]
47. Zhong, H.; Kim, S.; Zhi, D.; Cui, X. Predicting gene expression using DNA methylation in three human populations. *PeerJ* **2019**, *7*, 6757. [CrossRef]
48. Zhang, G.P.; Patuwo, B.E.; Hu, M.Y. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Comput. Oper. Res.* **2001**, *28*, 381–396. [CrossRef]
49. Liu, S.; Xu, M.; Wang, J.; Lu, F.; Zhang, W.; Tian, H.; Chang, G. A multilevel artificial neural network nonlinear equalizer for millimetre-wave mobile fronthaul systems. *J. Light. Technol.* **2017**, *35*, 4406–4417. [CrossRef]
50. Cong, S.; Liang, Y. PID-like neural network nonlinear adaptive control for uncertain multivariable motion control systems. *IEEE Trans. Ind. Electron.* **2009**, *56*, 3872–3879. [CrossRef]
51. Wang, H.Y.; Chang, S.C.; Lin, W.Y.; Chen, C.H.; Chiang, S.H.; Huang, K.Y.; Chu, B.Y.; Lu, J.J.; Lee, T.Y. Machine Learning-Based Method for Obesity Risk Evaluation Using Single-Nucleotide Polymorphisms Derived from Next-Generation Sequencing. *J. Comput. Biol.* **2018**, *25*, 1347–1360. [CrossRef]
52. Tu, J.V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epidemiol.* **1996**, *49*, 1225–1231. [CrossRef] [PubMed]
53. Alfonso Perez, G.; Castillo, R. Identification of Systemic Sclerosis through Machine Learning Algorithms and Gene Expression. *Mathematics* **2022**, *10*, 4632. [CrossRef]
54. Puleston, D.J.; Buck, M.D.; Klein, G.R.I.; Kyle, R.L.; Caputa, G.; O'Sullivan, D.; Cameron, A.M.; Castoldi, A.; Musa, Y.; Kabat, A.M.; et al. Polyamines and eIF5A Hypusination Modulate Mitochondrial Respiration and Macrophage Activation. *Cell Metab.* **2019**, *30*, 352–363. [CrossRef] [PubMed]
55. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **2019**, *14*, e0224365. [CrossRef] [PubMed]
56. McDonald, G.C. Ridge regression. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 93–100. [CrossRef]
57. Marquardt, D.W.; Snee, R.D. Ridge regression in practice. *Am. Stat.* **1975**, *29*, 3–20.
58. Hoerl, A.E.; Kannard, R.W.; Baldwin, K.F. Ridge regression: Some simulations. *Commun.-Stat.-Theory Methods* **1975**, *4*, 105–123. [CrossRef]