Predicting future cancer burden in the United States by artificial neural networks

Article

Predicting future cancer burden in the United States by artificial neural networks

Francesco Piva[1+], Francesca Tartari[2+], Matteo Giulietti[1], Marco Maria Aiello[3], Liang Cheng[4], Antonio Lopez-Beltran[5], Roberta Mazzucchelli[6], Alessia Cimadamore[6], Roy Cerqueti[2], Nicola Battelli[7], Rodolfo Montironi[6], and Matteo Santoni[7*]

[1] Department of Specialistic Clinical and Odontostomatological Sciences, Polytechnic University of Marche, 60126 Ancona, Italy; f.piva@univpm.it (FP); m.giulietti@univpm.it (MG);
[2] Department of Economics and Law, University of Macerata, via Crescimbeni, 20, 62100, Macerata, Italy; tartari.francy@gmail.com (FT); roy.cerqueti@unimc.it (RC);
[3] Oncology Unit, Policlinico Hospital, 95123 Catania, Italy; marcoaiello85@yahoo.it (MMA);
[4] Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA; liang_cheng@yahoo.com (LC);
[5] Department of Surgery, Cordoba University Medical School, 14071 Cordoba, Spain; em1lobea@gmail.com (ALB);
[6] Section of Pathological Anatomy, Polytechnic University of the Marche Region, School of Medicine, United Hospitals, 60126 Ancona, Italy; r.mazzucchelli@univpm.it (RMa – Roberta Mazzucchelli); alessiacimadamore@gmail.com (AC); r.montironi@univpm.it (RM – Rodolfo Montironi).
[7] Oncology Unit, Macerata Hospital, 62012 Macerata, Italy; mattymo@alice.it (MS); nicola.battelli@sanita.marche.it (NB);

++ These authors contributed equally to this work

** Correspondence: mattymo@alice.it; Tel: +39-071-5964830

Received: date; Accepted: date; Published: date

Abstract: The prediction of future cancer burden represents a key element for the design of cancer control programs and for the allocation of economic resources in the health system. To capture the complex relationships between risk factors and cancer incidences in the United States, we adopted an artificial neural network (ANN) algorithm. Data on the incidence of the four most common tumors (breast, colorectal, lung and prostate) from 1992 to 2013 (available from National Cancer Institute online datasets) were used for training and validation and data until 2050 were predicted.

According to our predictions, the rapid decreasing trend of prostate cancer incidence started in 2010 will continue until 2018/2019 and then it will slow down and reach a plateau after 2050, with several differences among different ethnicities. The incidence of breast cancer will reach a plateau in 2030, while colorectal cancer incidence will reach a minimum value of 30/100.000 in 2025, followed by a plateau in 2030 in men and in 2050 in women. As for lung cancer, the incidence will decrease from 53/100.000 (2015) to 42/100.000 in 2030 and 32/100.000 in 2050.

This up-to-date prediction of cancer burden in the United States could be a crucial resource for planning and evaluation of cancer-control programs. Urgent global actions towards a dramatic reduction of cancer-related risk-factors are actually needed and will accelerate the drop of incidences and the route to cancer eradication in future years.

Keywords: Artificial Neural Network; Breast cancer; Colorectal cancer; Future tumor burden; Lung cancer; Prostate Cancer.

1. Introduction

Cancer represents the second most lethal disease condition in the US, with more than 1,600 deaths per day and an estimated total that will exceed 600,000 Americans in 2018 [1]. The number of new diagnoses has progressively increased in the last decades getting over 1,700,000 new cases estimated in the US in 2018 [1]. However, it is important to note that in the last 25 years the rate of cancer-related deaths has declined for the four types of tumor with the highest incidences (prostate, breast, colorectal and lung) [1]. This event can be explained through both (1) the advances obtained by introduction of molecularly targeted drugs and novel immunotherapies as well as (2) the development of more effective diagnostic techniques and (3) the downward trend of cancer-related risk factors, above all smoking attitude.

The predictions of incidence cancer rates could be very useful to optimize the allocation of finite resources, the key elements of cancer control in next years and the future planning of cancer control programs. This has become rapidly fundamental due to the rising costs of oncologic treatments approved by the US Food and Drug Administration (FDA) in the last decades [2−5]. The trends of population growth and the ageing represent crucial factors for predicting the future burden of cancer, as the majority of tumors is age-dependent [6]. Furthermore, we might attempt to estimate the future incidence of different tumor types basing on the plausible changes of the risk patterns over time.

Previously, several prediction techniques have been employed to estimate future cancer trends. These methods include: linear extrapolation of trends [7,8], simple linear Poisson models [9,10] and classical [6, 11] or Bayesian age-period-cohort models [12]. However, the limits of these techniques [13] have led to the research of more accurate prediction models. To capture the complex relationships between input and target variables, we adopted an artificial neural network (ANN) algorithm. This mathematical model imitates the human brain strategy to solve problems and is able to extract knowledge directly from the raw data. These non-parametric modeling algorithms are very flexible and can perform any complex function mapping. Since ANN algorithms can train themselves under various circumstances, they are used in various fields, from finance to medicine [14-17]. In this study, we estimated the incidences of the four most frequent cancer diseases (prostate, breast, colorectal and lung) by developing an ANN algorithm to predict the number of new cases in the US till 2050.

2. **Materials and methods**

2.1 Data sets

USA population and USA life expectancy data derived from Gapminder (www.gapminder.org). Observed and projected obesity data in USA for male and female come from work of Wang et al. [18]. Number of cancer cases from 1975 to 2013 in USA was obtained by National Cancer Institute (https://progressreport.cancer.gov/diagnosis/incidence). We extracted data about tobacco consumption in developed countries derived from the work of Ng et al. [19]. Note that the best fitting polynomial to predict and interpolate missing data was Y = -0.3737 * X^2 – 3.7956 * X + 2363, where X is the year and Y the tobacco consumption.

2.2 Implementation of Artificial Neural Networks

Four ANNs, one for each cancer type, were built. To predict future incidence of prostate cancer in USA, we considered the three main risk factors internationally recognized: (1) demographic trend, (2) life expectancy data and (3) race/ethnicity [20−22]. Since the output also depends on race/ethnicity, we retrained this ANN with prostate cancer cases of White, Black, Hispanic, Asian/Pacific Islander and American Indian/Alaska Native races. Our predictions did not include prostate cancer with positive family history due to the lack of historical series in this setting. Data from 1992 to 2013 were used for training and validation and data from 2014 to 2050 were predicted.

Age is a commonly known factor associated with the risk of breast cancer [23]. It is estimated that females who survive till 85 years will present a lifetime rate of approximately 11% of developing breast cancer [23]. Similarly, obesity is associated with an increased risk of both premenopausal [24] and postmenopausal breast cancer [25], being also associated with worse cancer-related survival in

all breast cancer subtypes. To predict the incidence of breast cancer in USA, our ANN was based on (1) demographic trend, (2) life expectancy data and (3) obesity. Data from 1992 to 2013 were used for training and validation and data from 2014 to 2050 were predicted. As for prostate cancer, here we considered total population as an input variable.

The risk of developing colorectal cancer increases with age, with more than 90% of cases occurred in patients who are 50 years old or older [26]. It has been estimated that 29.5% of colorectal cancers can be attributable to a Body Mass Index (BMI)>22.5 [27], with an incidence that results different between men and women [28]. To predict the future incidence of colorectal cancer in USA, we included (1) demographic trend, (2) life expectancy data and (3) obesity as inputs of the ANN algorithm. Data from 1975 to 2013 were used for training and validation and data from 2014 to 2050 were predicted.

The vast majority of lung cancers can be attributable to tobacco attitude, with over 80% of lung tumors related to smoking [29]. As an indirect measure of exposure to pulmonary carcinogens (ionizing radiation, radon gas, etc), age can be considered as a predictive factor of lung cancer in both smokers and never-smokers patients [30]. Indeed, the relationship between age and lung cancer in non-smokers has been widely reported [31−33]. To predict the future incidence of lung cancer in USA, we considered (1) demographic trend, (2) life expectancy data and (3) smoking prevalence as input variables. Data from 1975 to 2013 were used for training and validation and data from 2014 to 2050 were predicted.

As a forecasting model we used multilayered perceptrons (MLP), a layered feed forward network that is trained by back propagation algorithm. Generally, a perceptron has several inputs and one output and the function connecting inputs and output is nonlinear. A back propagation training algorithm adjusts the connection strength between adjacent nodes. This method is easy to use, and it can model any kind of data although for training it takes longer time than other methods and it requires large amounts of training data. The flow of the algorithm that we used in this paper is showed in FigureS1. In order to obtain the best scores, the network structure and its internal parameters, hidden neuron number and transfer functions, the learning rate (LR) and learning momentum (LM) have been determined by many trials according to the trial-and-error method. Performance of each topology was assessed by the Mean Square Error and the Regression values. We sought to reach the best results by minimum number of nodes to avoid that the ANN memorized data rather learning them for generalization. Our ANNs are constructed with three layers, the input layer has as many neurons as inputs and the output layer has one neuron. The ANNs for colorectal, lung, breast and prostate cancers have respectively 10, 25, 20 and 20 neurons that implement the "tansig" function in the hidden layer. Software package Matlab R2014b (Mathworks Inc.) was used in this study. Inputs were scaled, by "mapminmax" function of Matlab, to fall in between -1 and 1 in order to account for different variations and degrees of magnitude of input variables. We used 70% of the data for training and 30% for validation and these data sets were chosen randomly.

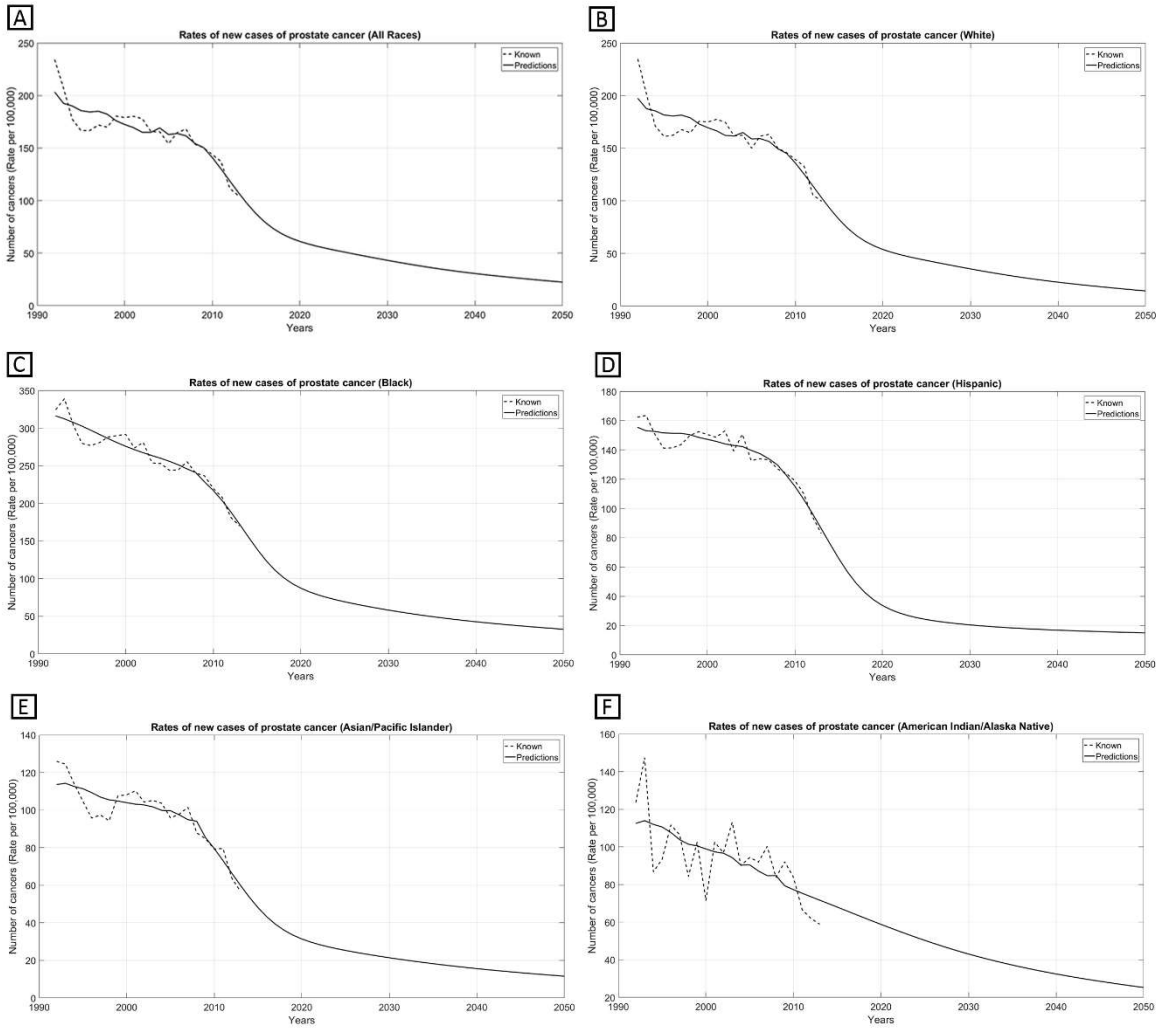
3. Results

3.1 Prostate cancer

The incidence of prostate cancer has decreased from an incidence of 200 cases/100.000 habitants registered in the 1990s in USA to less than 150/100.000 in 2010 and, according to our predictions, will successively fall under 50/100.000 from 2025 (Figure 1A). As for all the different races/ethnicities, our algorithm predicts that the rapid decreasing trend started in 2010 will continue until 2018/2019 and then it will slow down and will reach a plateau after 2050 (Figure 1A). The trend observed in the overall population reflects that reported in white patients, characterized by an incidence of less than 200/100.000 cases in the 1990s, with a drop under 50/100.000 in 2020s (Figure 1B). The incidence is superior in black patients (Figure 1C), who have been associated with an incidence lower that 200/100.000 only from 2012 and will reach a value of less than 50/100.000 only more than ten years later (Figure 1C). Otherwise, patients with Asian/Pacific ethnicity as well as American Indian and

Alaska native are characterized by a lower incidence (Figure 1E, 1F). Only for American Indian/Alaska Native Races the decreasing trend is almost steady (Figure 1F).

Figure 1. Trend and predicted new cases of prostate cancer overall (A) and by ethnicity (B=White; C=Black; D=Hispanic; E=Asian/Pacific Islander; F=American Indian/Alaska Native).

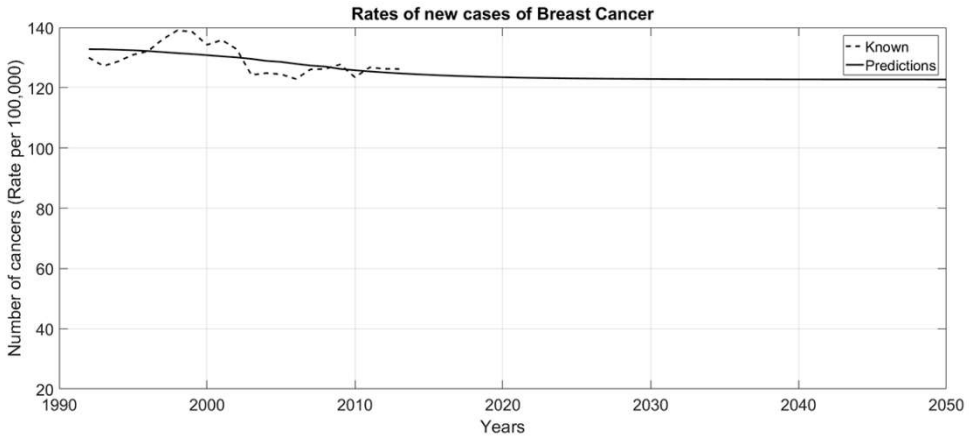


The fading in prostate cancer decreasing from 2018 could reflect the fading in life expectancy and population increasing. The racial disparities are caused by behavioral differences and unequal access to high-quality health care but this gap is rapidly reducing.

3.2 Breast cancer

The incidence of breast cancer has slightly decreased from 1990s (133 cases/100.000 habitants), registering a drop to 125/100.000 in 2002 and to 124/100.000 in 2015 (Figure 2). Based on our prediction algorithm, the incidence will decrease to 123/100.000 in 2020, reaching a plateau in 2030 (Figure 2). Performance of Train and Validation phases (Figure S3) showed that, in this setting, the ANN algorithm obtained worst predicting results (Performance of Train = 0.641; Validation phases = 0.577) compared to prostate, colorectal and lung cancer. The reasons of this different behavior are discussed in the next sections.

Figure 2. Trend and predicted new cases of breast cancer. Our calculations are based on population, life expectancy and obesity data for female.

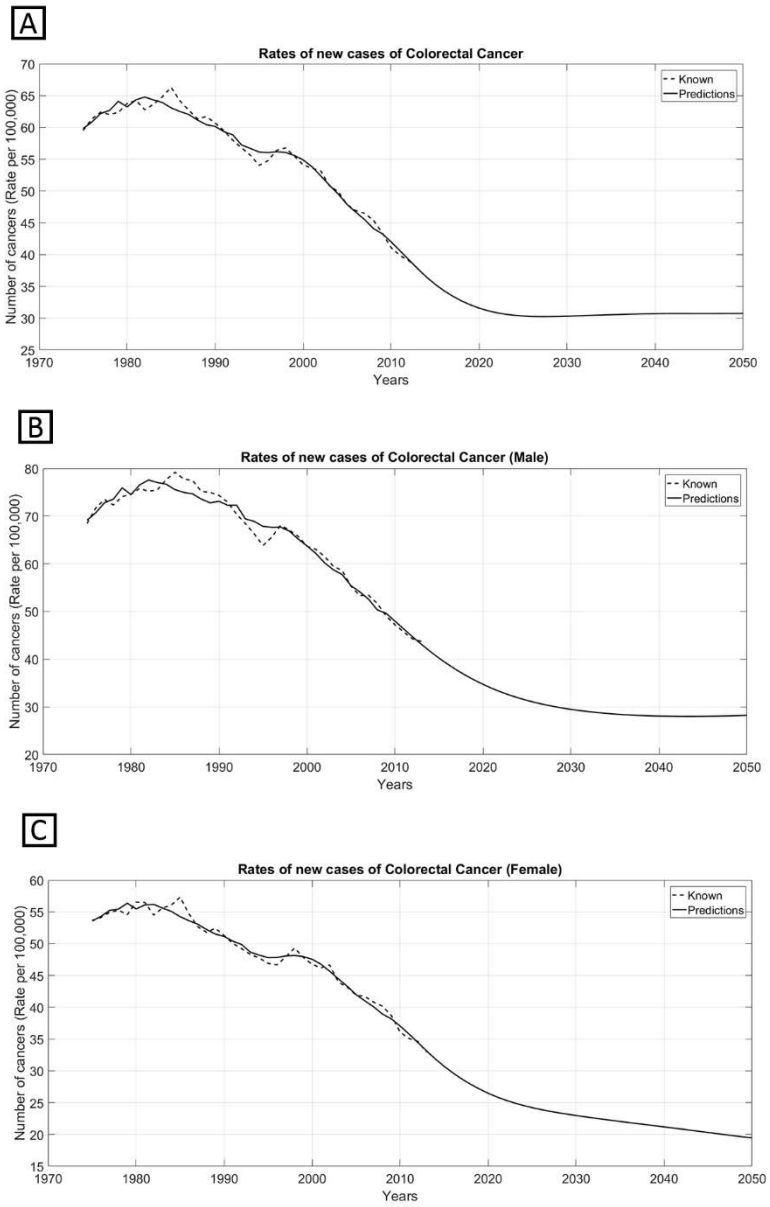


3.3 Colorectal cancer

The incidence of colorectal cancer has progressively increased from 1970s (60 cases/100.000 habitants), registering its maximum in 1985 (66/100.000) (Figure 3A). From late 1980s, the downward trend led an incidence of 55/100.000 in 2000, with a progressive reduction to 35/100.000 in 2015 (Figure 3A). Based on our prediction algorithm, the incidence will reach a minimum value of 30/100.000 in 2025, followed by a plateau until 2050 (Figure 3A).

Due to the influence of gender on colorectal cancer incidence [33], we further predicted the incidence in males (Figure 3B) and females (Figure 3C). We found that the incidence was maximum for men in 1985 (79/100.000) and for women in 1985 (57/100.000), respectively (Figure 3B, 3C). Interestingly, we predicted that the reduction of the incidence trend will lead to a plateau around 30/100.000 in 2030 in men, while women will be associated with a lower incidence that will drop below 20/100.000 in 2050 (Figure 3B, 3C).

Figure 3.Trend and predicted new cases of colon cancer overall (A) and by gender (B=males; C=females). Our calculations are based on population, life expectancy and obesity for males and females.

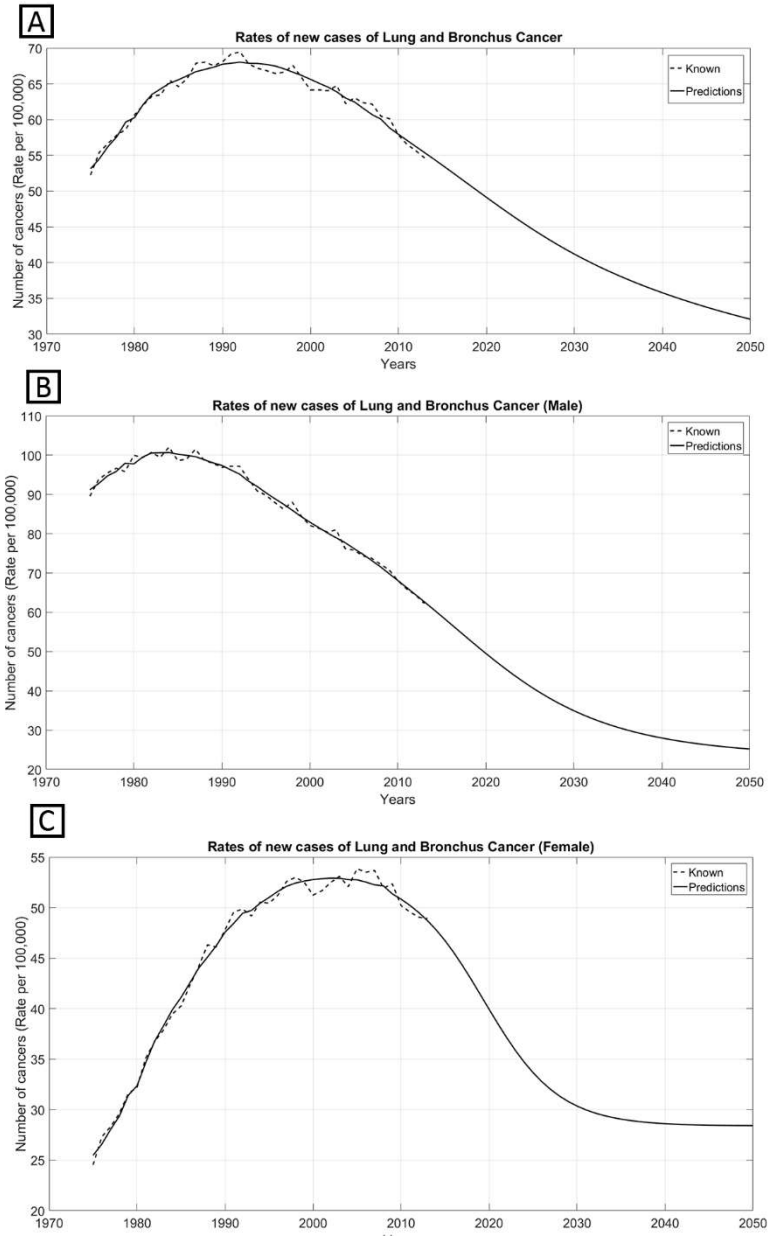


3.4 Lung cancer

The incidence of lung cancer has progressively increased from 1970s (53 cases/100.000 habitants), registering its maximum in 1992, characterized by 69/100.000 (Figure 4A). From 1990s, the downward trend in the smoking attitude has led to a gradually reduction of new cases of lung cancer, with an incidence in 2015 comparable to 1970s (53/100.000) (Figure 4A). Based on our prediction algorithm, the incidence will decrease to 42/100.000 in 2030 and fall to 32/100.000 in 2050 (Figure 4A).

Due to the different time-trends in tobacco consumption, we further predicted the incidence in males (Figure 4B) and females (Figure 4C). While the maximum incidence was registered in 1984 for males (102/100.000), the highest value (54/100.000) for women was reported in 2005 due to the rapid increase of smoking prevalence among women about 20 years later than men (Figure 4B, 4C). Interestingly, the drop of the incidence trend appears slower in males than in females, reaching a plateau beyond 2050 (25/100.000), 15 years after the plateau predicted for females (28/100.000) (Figure 4B, 4C).

Figure 4. Trend and predicted new cases of lung cancer overall (A) and by gender (B=males; C=females). Our calculations are based on population, life expectancy and smoking data for males and females.



3.5 Regression Analysis

The ANN outputs with respect to targets for training and validation sets are shown in scatter plots (Figures S2-S5). The dashed line represents the perfect result, i.e.: outputs = targets. The solid line depicts linear best fit between the outputs and the targets. The R-value summarizes the relationship between the outputs and targets. In particular, if R=1 there is an exact linear relationship between outputs and targets. If R is close to zero means that a no linear relationship links outputs to targets. In almost all our models, apart for breast cancer and prostate cancer in American Indian/Alaska Native Races, the R-values around 0.9 or greater indicate a reasonably good fit for a data set. Analogously, a very small value of mean square error (MSE) suggests the goodness of the models.

4. Discussion

The lifetime risk of developing cancer is related, in general, to a longer life expectancy. However, a range of influences, from environmental and attitude changes to prevention campaigns, screening programs and innovation technologies, should be taken into account in order to increase the accuracy of predicting models for cancer incidence.

Among the four most frequent tumor types we showed a general decline in the incidences in the United States. The changes in prostate cancer incidence observed in the late 1980s and early 1990s (Figure 1A) were probably due to the introduction of widespread prostate-specific antigen (PSA) testing that allowed the detection of asymptomatic disease [34]. The reduction in prostate cancer incidence from 2010 to 2013 can be attributed to decreased PSA testing. In fact, the US Preventive Services Task Force (USPSTF) diffused a recommendation about the use of PSA as a screening method for prostate cancer. The task force, basing on data from Prostate, Lung, Colorectal and Ovary cancer screening study (PLCO) and the European Randomized Study of Screening for Prostate Cancer (ERSPC) trial, informed that the potential harms of testing (erectile dysfunction, incontinence and serious surgical complications) overcame the benefits (PSA screening reduced cancer-related mortality by 4 men for every 1000 men, after 14 years of follow-up) [35].

In breast cancer, the ANN did not reach good performances. The small variability of incidence data from 1990 to 2050, together with the high multifactoriality of this tumor, may partially explain the less good performance of our ANN in this disease (as evident from the regression curves in Figure S3). This evidence suggests that the number of new cases does not strictly follow the trends registered for age and obesity in USA (i.e. the peak from 1995 to 2002 in cancer incidence in a time-interval characterized by the reduction of both risk factors) and underlines the necessity of identifying more effective input variables beyond the most commonly recognized risk factors.

As for colorectal cancer, the drop of the incidence rates before 2000 should be explained by the changes in risk factors and the introduction of screening (Fecal Occult Blood Testing (FOBT) and endoscopy) [36]. The prevalence results distinct between men and women due to a series of underlying different mechanisms that include estrogen exposure, menopausal status, insulin resistance, chronic inflammation and steroid hormones [37, 38].

Lung cancer is among the most deadly cancers for both men and women [39]. Reducing its incidence represents a major goal for cancer researchers, and both the results of these enforces and the worldwide prevention campaigns to decrease tobacco consumption find a mirror in Figure 4A showing the falling incidence of lung cancer. This progressive reduction would be even more rapid as an effect of the global action towards the 2040 tobacco-free world goal [40]. As for the gender differences (Figure 4B, 4C), they reflect the historical attitudes in tobacco use, with women starting to smoke in large numbers later and at older ages than men.

Our study presents several limitations. As other prediction systems, ANN algorithms are affected by errors and biases compared to real data. However, ANNs provide for training, performance and validation phases that may partially reduce system biases and increase the accuracy of predictions.

5. Conclusions

This up-to-date prediction of cancer burden in the United States could be a crucial resource for planning and evaluation of cancer-control programs. Urgent global actions towards a dramatic reduction of cancer-related risk-factors are actually needed and will accelerate the drop of incidences and the route to cancer eradication in future years.

Supplementary material:

Figure S1. The flow of the Artificial Neural Network (ANN) algorithm employed in this study.

Figure S2. Performance of Train and Validation phases for prostate cancer by ethnicity. Data were scaled into the range used by the input neurons in the neural network. In this case the range is -1 to 1 (A=all races; B=White; C=Black; D=Hispanic; E=Asian/Pacific Islander; F=American Indian/Alaska Native).

Figure S3. Performance of Train and Validation phases for breast cancer. Data were scaled into the range used by the input neurons in the neural network. In this case the range is -1 to 1. Mean squared normalized errors for the Train and Validation sets are respectively 0.27882 and 0.14008.

Figure S4. Performance of Train and Validation phases for breast cancer. Data were scaled into the range used by the input neurons in the neural network. In this case the range is -1 to 1 (A=total population; B=males; C=females).

Figure S5. Performance of Train and Validation phases for lung cancer. Data were scaled into the range used by the input neurons in the neural network. In this case the range is -1 to 1 (A=total population; B=males; C=females).

Author contribution: conceptualization: MS and FT; methodology: FP, MG, MMA; data curation: FP, MS, RM (Roberta Mazzucchelli), RC; writing (original draft preparation): FT, MS, FP, NB; writing (final editing): LC, ALB, AC, RM (Rodolfo Montironi).

Funding

The research receive no external funding.

Conflicts of Interest

The authors declare to have no conflicts of interest.

Abbreviations

ANN	Artificial Neural Network
BMI	Body Mass Index
FDA	Food and Drug Administration
FOBT	Fecal Occult Blood Testing
LM	Learning Momentum
LR	Learning Rate
MLP	Multilayered Perceptrons
MSE	Mean Square Error
PSA	Prostate-Specific Antigen
USPSTF	US Preventive Services Task Force

References

1. http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2018; [accessed 22th August 2019].
2. Tartari F, Santoni M, Burattini L, Mazzanti P, Onofri A, Berardi R. Economic sustainability of anti-PD-1 agents nivolumab and pembrolizumab in cancer patients: Recent insights and future challenge. *Cancer Treat Rev*. **2016**;*48*:20-4. doi: 10.1016/j.ctrv.2016.06.002.
3. Montironi R, Santoni M, Tartari F, Lopez-Beltran A, Cheng L, Berardi R, Scarpelli M. Testing PD-1/PD-L1 Expression in Cancer Therapy: Pathologic Insights and Economic Sustainability. *Arch Pathol Lab Med*. **2016**;*140*(6):501-2. doi: 10.5858/arpa.2015-0529-LE.
4. Tartari F, Santoni M, Pistelli M, Berardi R. Healthcare cost of HER2-positive and negative breast tumors in the United States (2012-2035). *Cancer Treat Rev*. **2017**;*60*:12-17. doi: 10.1016/j.ctrv.2017.08.005.
5. Tartari F, Conti A, Cerqueti R. Assessing the relationship between toxicity and economic cost of oncological target agents: A systematic review of clinical trials. *PLoS One*. **2017**;*12*(8):e0183639. doi: 10.1371/journal.pone.0183639.
6. Moller, B. et al. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur. J. Cancer Prev.* **2002**; *11 (Suppl.)*, S1–S96.
7. Teppo, L., Hakulinen, T. & Saxen, E. The prediction of cancer incidence in Finland for the year 1980 by means of cancer registry material. *Ann. Clin. Res*. **1974**; *6*, 122–125.
8. Hakulinen T, Teppo L. & Saxen, E. Do the predictions for cancer incidence come true? Experience from Finland. *Cancer* **1986**; *57*, 2454–2458.
9. Hakulinen, T & Dyba, T. Precision of incidence predictions based on poisson distributed observations. *Statist. Med.* **1994**; *13*, 1513–1523.

10. Dyba, T., Hakulinen, T. & Paivarinta L. A simple non- linear model in incidence prediction. *Statist. Med.***1997**; *16*, 2297–2309.
11. Engeland, A. et al. Prediction of cancer incidence in the Nordic countries up to the years 2000 and 2010. A collaborative study of the five Nordic Cancer Registries. *APMIS* **1993**; *101 (suppl.),* 5–123.
12. Coory, M. & Armstrong, B. K. Cancer Incidence Projections for Area and Rural Health Services in New South Wales. *NSW Cancer Council, Sydney*, **1998**.
13. Bray F, Møller B. Predicting the futureburden of cancer. *Nat Rev Cancer.* **2006**;*6*(1):63-74.
14. Malik F, Nasereddin M. Forecasting output using oil prices: A cascaded artificial neural network approach. Journal of Economics and Business. 2006; 58(2):168–80. http://dx.doi.org/10.1016/j.
15. Kim D. Neural networks theory and applications. Seoul: Jinhan M&B; 2005.
16. Cooper JCB. Artificial neural networks versus multivariate statistics: An application from economics. *Journal of Applied Statistics*. **1999**; *26*(8):909–21. doi: 10.1080/02664769921927
17. Zhang G, Eddy Patuwo B, Y Hu M. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting.* **1998**; *14*(1):35–62.
18. Wang Y, Beydoun MA, Liang L, Caballero B, Kumanyika SK. Will all Americans become overweight or obese? estimating the progression and cost of the US obesity epidemic. *Obesity (Silver Spring).* **2008**; *16*:2323-2330.
19. Ng M, Freeman MK, Fleming TD, Robinson M, Dwyer-Lindgren L, Thomson B, Wollum A, Sanman E, Wulf S, Lopez AD, Murray CJ, Gakidou E. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. *JAMA*. **2014**; *311*:183-192.
20. Gann PH. Risk factors for prostate cancer. *Rev Urol.* **2002**;*4 Suppl* 5:S3-S10.
21. Leitzmann MF, Rohrmann S. Risk factors for the onset of prostaticcancer: age, location, and behavioral correlates. *Clin Epidemiol*. **2012**;*4*:1-11.
22. Kheirandish P and Chinegwundoh F. Ethnic differences in prostate cancer. *Br J Cancer* **2011**; *105(4*): 481–485.
23. Singletary SE. Rating the Risk Factors for Breast Cancer. *Ann Surg* **2003**; *237*(4): 474–482.
24. Laudisio D, Muscogiuri G, Barrea L, Savastano S, Colao A. Obesity and breast cancer in premenopausal women: Current evidence and future perspectives. *Eur J Obstet Gynecol Reprod Biol*. **2018**; pii: *S0301-2115*(18)30148-9.
25. Argolo DF, Hudis CA, Iyengar NM. The Impact of Obesity on Breast Cancer. *Curr Oncol Rep* **2018**;*20*(6):47.
26. Farin Amersi, Michelle Agustin, Clifford Y. Ko. Colorectal Cancer: Epidemiology, Risk Factors, and Health Services. *Clin Colon Rectal Surg*. **2005**;*18*(3): 133–140.
27. Larsson SC, Bergkvist L, Wolk A. High-fat dairy food and conjugated linoleic acid intakes in relation to colorectal cancer incidence in the Swedish Mammography Cohort. *Am J Clin Nutr.* **2005**;*82*:894–900.
28. Kim SE, Shim KN, Jung SA, Yoo K, Moon IH. An association between obesity and the prevalence of colonic adenoma according to age and gender. *J Gastroenterol.* **2007**;*42*:616–23.
29. Centers for Disease Control and Prevention (CDC) Annual smoking-attributable mortality, years of potential life lost, and productivity losses--United States 1997–2001. *MMWR Morb. Mortal. Wkly. Rep.* **2005**;*54*(25):625–628.
30. Doll R. The age distribution of cancer: implications for models of carcinogenesis. *Journal of the Royal Statistical Society.Series A (General)* **1971**;*134*(2):133–166.
31. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J. Epidemiol. Community Health.* **1978**;*32*(4):303–313.
32. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model. *Cancer Causes Control*. **2008**;*19*(3):317–328.
33. Thun MJ, Henley SJ, Burns D, Jemal A, Shanks TG, Calle EE. Lung cancer death rates in lifelong nonsmokers. *J. Natl. Cancer Inst.* **2006**;*98*(10):691–699.
34. Potosky AL, Miller BA, Albertsen PC, Kramer BS. The role of increasing detection in the rising incidence of prostate cancer. *JAMA.* **1995**; *273*:548-552.
35. Moyer VA; U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med.* **2012**; *157*:120-134.

36. Edwards BK, Ward E, Kohler BA, Eheman C, Zauber AG, Anderson RN, Jemal A, Schymura MJ, Lansdorp-Vogelaar I, Seeff LC, van Ballegooijen M, Goede SL, Ries LA. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer*. **2010**; *116*:544-573.
37. Slattery ML, Ballard-Barbash R, Edwards S, Caan BJ, Potter JD. Body mass index and colon cancer: an evaluation of the modifying effects of estrogen (United States). *Cancer Causes Control*. **2003**; *14*(1):75-84.
38. Jochem C, Leitzmann M. Obesity and Colorectal Cancer. Recent Results Cancer Res. 2016;208:17-41.
39. Siegel R, Ma J, Zou Z, et al. Cancer statistics, 2014. *CA Cancer J Clin* **2014**;*64*:9-29.
40. Beaglehole R, Bonita R, Yach D, Mackay J, Reddy KS.nA tobacco-free world: a call to action to phase out the sale of tobacco products by 2040. *Lancet* **2015**;*385*(9972):1011-8.