Machine Learning as a Next-Generation Tool for Indoor Air Radon Exposure Prediction

Contributors: Selim M. Khan, Joshua M. Taron & Aaron A. Goodarzi
Pub. Date: 2020
Product: SAGE Research Methods Cases: Medicine and Health
Methods: Data visualization, Machine learning, Prediction
Disciplines: Public Health
Access Date: July 31, 2020
Academic Level: Introductory Undergraduate
Publishing Company: SAGE Publications Ltd
City: London
Online ISBN: 9781529743708
DOI: https://dx.doi.org/10.4135/9781529743708

Abstract

Indoor air quality is strongly influenced by the presence of radioactive radon (^{222}Rn) gas. Indeed, exposure to high ^{222}Rn concentrations is unequivocally linked to DNA damage, lung cancer, and is a worsening issue in North American residences, having increased over time within newer housing stocks as a function of as yet unclear variables. Indoor air radon concentration can be influenced by a wide range of environmental, structural, and behavioral factors. As some of these factors are quantitative while others are qualitative, no single statistical model can determine indoor radon level precisely, while simultaneously considering all these variables across a complex and highly diverse dataset. The ability of machine learning to simultaneously analyze multiple quantitative and qualitative features makes it suitable to predict radon with a high degree of precision. Using Canadian long-term indoor air radon exposure data, we are using an artificial neural network with random weights and polynomial statistical models in MATLAB to predict radon levels as a function of geospatial and built environmental metrics. Our initial artificial neural network with random weights model run by sigmoid activation tested different combinations of variables and showed the highest prediction accuracy within the reasonable iterations. Here, we present details of these emerging methods and discuss strengths and weaknesses compared to the traditional artificial neural network and statistical methods commonly used to predict indoor air quality in different countries. We propose artificial neural network with random weights as a highly effective method for predicting indoor radon.

Learning Outcomes

By the end of this case, students should be able to

- Understand and describe the applications for artificial neural network with random weights (ANNRW) and polynomial regression models in the context of indoor air quality analysis
- Compare and contrast the strengths and weaknesses in predicting indoor radon using these methods
- Evaluate and appraise the efficiency of ANNRW versus other models to produce robust outcomes when assessing indoor air contaminants

Project Overview and Context

Understanding indoor air quality (IAQ) is crucial for the health protection of people living in cold countries, as they spend approximately 90% of their time indoors, working, studying, playing, or sleeping—a phenomenon often driven by increased sedentary behavior and/or long heating or air conditioning seasons (Klepeis et al., 2001; Setton et al., 2013). The short-lived radioactive gas radon (^{222}Rn) is a prevalent IAQ pollutant across the world and is particularly abundant within the built environment of the North American Prairies (Gaskin et al., 2018; Stanley et al., 2017, 2019). Decaying ^{222}Rn emits alpha particle ionizing radiation, a potent cancer-causing agent that, even in doses observed commonly within the residential built environment, causes highly mutagenic DNA damage that drives human and animal lung cells toward becoming cancer (Darby et al., 2005; Krewski et al., 2005; Moore et al., 2014; Pearson et al., 2016). With some of the world's most abundant reserves of high-grade uranium (^{238}U) and a surficial geology enriched for other radon-source materials (including radium and thorium), Canada has an exceptionally large geologic radon potential (Barnett & Bajc, 2002; Garrity & Soller, 2009). A recent study found that residential buildings in the North American Prairie region of Canada contain, on average, the second highest levels of radon observed globally to date (Gaskin et al., 2018; Stanley et al., 2017, 2019). Perhaps unsurprisingly, these areas also experience high rates of never-smoker lung cancer, with approximately one in five of all lung cancer patients not being tobacco users, and an estimated 10,000–40,000 new radon-attributable lung cancers being diagnosed per decade (Grundy et al., 2017; Health Canada, 2012; World Health Organization, 2009). Thus, this is a very important problem with the substantial burdens of human misery and health economic impact yet is preventable with a complete understanding of how the exposure occurs.

When constructed on sites with high radon potential, airtight buildings with poor ventilation can create conditions for hazardous radon gas exposure (World Health Organization, 2009). Radon entry from soils and rocks beneath foundations can occur through numerous vectors, including foundational cracks, foundation-to-wall gaps, non-airtight sump pumps (and other ground penetrations), and/or through water. Radon accumulation, once thought to always be a winter-heating month problem, is now increasingly recognized as a year-round or, for some homes, even a summer issue (Papastefanou et al., 1994; Stanley et al., 2019; Wilson et al., 1991). This latter point is thought to have increased with time, as more and more properties acquire air conditioning, and/or increasingly prevalent and severe forest/bush fires during summer preclude window opening behavior (and air ventilation) due to prolonged periods of toxic smoke pollution. As discussed, a variety of environmental, structural, and occupant behavioral factors determine indoor radon levels, and it is not possible to apply any single metric to predict radon with any degree of accuracy. Effective mitigation measures must consider all the potential factors that contribute to indoor radon. A previous study has compared traditional artificial neural networks (ANNs) with multiple nonlinear regression (MNLR) models in ambient radon prediction, finding that traditional ANNs showed enhanced accuracy relative to the MNLR model (Kulikajevas et al., 2019). This, however, is not the norm in terms of analyzing radon as an IAQ metric.

We aim to apply a next-generation machine learning analytics, specifically the artificial neural network with random weights (ANNRW) algorithm, in radon risk modeling to ensure maximum precision to generate evidence-based radon exposure solutions at the individualized level (i.e., mitigating a specific home) and, perhaps more importantly, policy and societal level (i.e., engineering out radon exposure by modifying build practices). Artificial deep neural networks (DNNs) have gained attention for accomplishing difficult tasks with greater precision in less time versus other techniques (Siuly & Zhang, 2020). The application has, perhaps, most usefully been applied in (a) speech, image, video and, object recognition; (b) remote diagnosis (i.e., radiology); (c) establishing drug effectiveness; (d) pattern recognition in big data; and (e) time series forecasting and more (Kulikajevas et al., 2019; Lee et al., 2017; Siuly & Zhang, 2020). Different machine learning algorithms have also been applied in IAQ assessment. However, ANN has not yet been applied for indoor radon prediction.

Section Summary

- IAQ is an important public health determinant.
- Predicting indoor radon has been a challenge as multiple factors influence exposures.
- Machine learning has shown potential to predict IAQ with precision.

Research Design

"Evict Radon" is the broad title for a series of interconnected public research projects (approved by the Research Ethics Boards of the University of Calgary (REB ID#17-2239) that have the specific goal of understanding Canadian radon exposure at all levels. It intersects public health, radiation biology, building science and architecture, geology, psychology, policy, and communications (see www.evictradon.org for more details). Project leads adhere to the regional guidelines and regulations for research involving citizen science participants. The survey region encompassed the Western North American Prairies (Alberta, Saskatchewan), a region with >300 Bq/kg (of surficial material) radon potential based on geologic analysis (Stanley et al., 2019). For context, 1 Bq (Becquerel) equates with one radioactive emission of particle radiation per second. Radon exposure is most typically measured in Bq per cubic meter of air (hence, Bq/m^3).

Evict Radon project participants are randomly targeted, and following informed consent, adult "citizen scientists" are provided secure links to purchase alpha track 90 + day radon detectors (at cost only, and quality controlled by researchers). Participants are guided through the deployment, registration, testing and return process, and complete online surveys to define >40 individual property metrics for the building being radon tested, as well as geospatial information. The cohort consented to provide data on the construction year, build type, foundation type, furnace type, heat-delivery type, floor tested, room of deployment, ceiling heights, thermostat settings, window opening behavior, basement and ground floor surface area (square footage), and thermostat settings. Questionnaires were completed online with dropdowns for defined response options; this was essential for statistical analysis. Data are otherwise de-identified and ultimately aggregated to preserve privacy of participants. Participants are not selected from lung cancer cohorts, and demographic analysis reveals a generally representative population in terms of sex, gender, age (considering only adults are eligible), and income. These projects are still ongoing and are now national in scope, collecting radon exposure linked to building metrics from across all areas of Canada. The study remains open to all eligible adult Canadians.

Section Summary

- Study followed approved protocols on design, methodology, and data collection.
- Selection process was random, open to both homeowners and tenants.
- Build metric questionnaires were constructed to generate data suitable for statistical analysis.

Research Practicalities

We took particular efforts to ensure the highest quality source data (Stanley et al., 2019) that is the topic of this case report. All radon testing kits were quality controlled and distributed centrally by researchers. Our research team took rigorous care to educate participants in the correct test deployment methods through communication with the Canadian National Radon Protection Program (C-NRPP). We strictly maintained adherence to the advice and testing protocols to Health Canada's guidelines. For long-term (90 + day) alpha track radon tests, participants were advised to place devices on the lowest level of the building occupied for approximately four or more hours per day, for a minimum of 3 months (during the typical Canadian heating season; October to April) or longer. Five percent of all participants were randomly selected for free duplicate devices, to be placed <10 cm apart from their primary devices to ensure data precision. The outcome was 96.2% precision (meaning two devices side by side predicted one another 96.2% of the time). We also performed accuracy controls, including blanks (nonexposed tests) and spiked positives (a subset of devices exposed to know amounts of radon), demonstrating >99% accuracy. We recorded the location of test kit placement of residential buildings to maintain consistency.

Although in the region, the majority of houses have basements or cellars, and they also use similar foundational materials, it was very difficult for anyone (participant or researcher) to reasonably predict a property to have "high" or "low" radon given the functionally uniform geologic radon potential across the entire survey region. Hence, the possibility of selection bias was fully considered and controlled. Ethical considerations regarding autonomy and confidentiality of personally identified information were carefully maintained to the fullest extent possible. A comprehensive discussion of the role of research ethics in citizen science for projects like this can be accessed here (Oberle et al., 2019), and full details of study design can be accessed here (Stanley et al., 2019).

Section Summary

- Rigorous care was taken to ensure correct test placement, duration, accuracy, and precision.
- Researchers remained vigilant to avoid or minimize any possible selection bias.
- Care was taken to meet ethics requirements.

Methods

Selection and Significance

Having scanned the literature for options for a more sophisticated approach to pattern recognition within complex data, we chose to explore the potential of artificial deep neural networks (ADNN) for analyzing our data. To date, machine learning algorithms have had diverse applications including in assessing IAQ. The question is how well can ADNN accomplish this task? ADNN is, essentially, "just" a network of artificial neurons, representing supervised machine learning algorithms of artificial intelligence that emulates the function of the human brain. A biological neuron has dendrites that receive sensory inputs from its surroundings, processes these inputs within cell body, and sends forward that process signal to axons that transmit the outputs. Basically, an action signal from one neuron to another propagates onwards to other cells and organs. ADNN actually works in a way different from traditional computer programming. Whereas a traditional computer-based algorithm would simply cease working in the event of a missing step in the instruction pathway, the ADNN has a self-learning capability and it can make predictions even if some of the provided features are missing. In such cases, it has the capability to decide based on the available information.

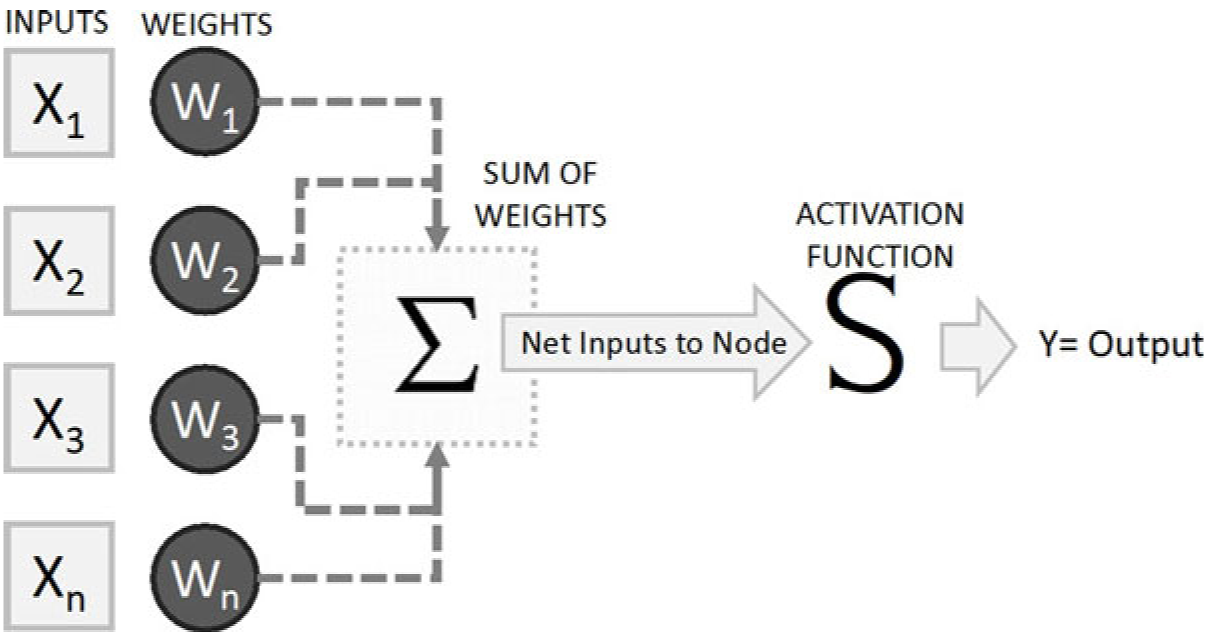
Steps and Metrics

The steps of ADNN start with reading the dataset, defining the features, labeling, and encoding the variables (Figures 1 and 2). Then, it is necessary for the analysts to train the ANN model. To accomplish this, we randomly divided the dataset into three parts, for example, training (70%), cross-validation (15%), and testing (15%). In our model, the accuracy performance metrics were mean square error; minimum, maximum, and mean absolute error (MSE, Min/MaxAE, and MAE); coefficient of determination (r^2) derived from all training, cross-validation, and independent validation of test data. We then compared them with the corresponding error terms and r^2 for multiple polynomial regression (MPNR) models performed in the same operating system.

Methods in Action

We trained a given model while also cross validating its accuracy. In this process, unimportant features were removed recursively, and under-fitting and over-fitting issues were resolved. In doing so, we ran the model in enough iterations with random modification of weights until MSE was reduced to the minimum possible level. Once we were satisfied with the performance, our model was then considered trained, and ready for implementation in making predictions on the test datasets. We developed and tested both ANNRW and MPNR models in MATLAB R2019a for academic use (specifically “The Math Works, Inc.” MA, USA. © 1994–2020).

Figure 1. Perceptron: single artificial neuron.



Source: Authors.

ANN in Detail

To describe ANN in more detail, but with a degree of simplicity, it is important to understand that one (distinct) artificial neuron is called a "perceptron" (Figure 1). A single perceptron has only one input and one output layer, but no hidden layer. This is a single-layer neural network that expresses a linear function. If a single hidden layer is added, this neural network is called a "shallow" or "vanilla" neural network. When two or more hidden layers are added, this is called deep neural network (DNN). In all multiple ANNs, there are more than one hidden layer and the function is nonlinear. A unique aspect to note here is that, whereas human brain uses associations of neurons at synapses to retain memories, ANNRW does so with the connection weights. Training of the traditional neural networks mostly depends on running back propagation in reducing error that gradually adjusts the parameters. Once the number of hidden layers increases, training data needs more time due to slow convergence; thus, instead of reaching global minimum, we can find many local minima. Global minimum is to find the lowest cost function (a function that measures the performance of a Machine Learning model for given data; see further readings, Macdonald, 2017) with gradual descent through an optimum learning rate whereas many local minima are produced when the learning rate cannot be set to an optimum. To evade these problems, we applied new-generation neural networks using random weights (ANNRW). This ANNRW assigns random weights between the input and hidden layers, whereas the weights between the last hidden layer and output layer are analytically calculated. Recent study has demonstrated that such random weight allocation reduces the complexity in training data compared to the traditional ANN training method (Cao et al., 2018).

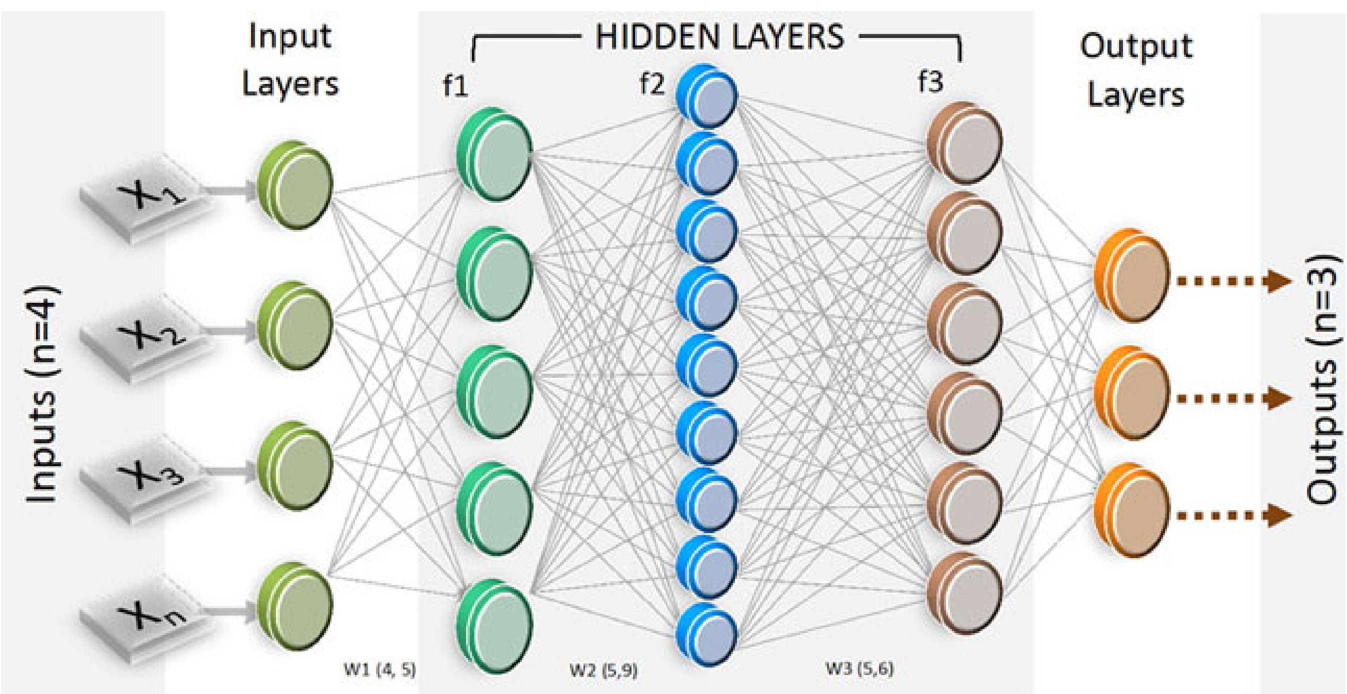
What Do Connection Weights Mean?

As Figure 1 shows, there are multiple inputs of ANN, for example, X_1, X_2, X_3, …, X_n (features or independent variables). At the first input level, the variables do not use weighted sum and do not need any activation function. These are similar to the signals that are received by the dendrites of a neuron. Each input of ANN is then assigned an adequate weight to initialize the network. The defined weights are marked as W_1, W_2, W_3, …, W_n. These weights are summed up and if the net amount reached to a particular threshold level, an activation function is generated at the next node. These mathematical calculations at the second level (cell body) then determine the behavior of the node, and (put simply) this is just the generation of a threshold that triggers a particular function at the next level of nodes. The process continues till it creates the outputs, that is, Y_1, Y_2, Y_3, …, Y_n (labels or dependent variables). The output is again the weighted sum of all the signal paths in the network plus an associated bias. Here, $Y = W \bullet X + b$ or $((W_1 \bullet X_1) + (W_2 \bullet X_2) + (W_3 \bullet X_3) \ldots (W_n \bullet x_n) + b)$. Therefore, it is evident that higher the weight, the greater will be the output. In case where the threshold is not generated at any level of nodes, there will be no output at all.

Phases and Modes

It is also notable that the perceptron has two phases: training phase and using phase. "Training phase" is when the neuron is trained to generate a threshold to trigger an action, and the "using phase" is when training is completed, and the model becomes ready to detect inputs and generate outputs. There are many modes of activation functions, and the most prominent ones being the step function, sign function, and sigmoid function. How this activation function works depends on factors that can generate particular threshold. These factors have assigned weights and, depending on that and the set threshold level, the neuron will trigger (or not).

Figure 2. Deep artificial neuronal networks.



Source: Authors.

New-Generation ANN

In our case, we ran the most powerful new-generation ANNRW (Figure 2) algorithm for supervised training in a multiple/deep neural network. The stepwise advancement of the activation functions is called feed-forward propagation. As the process creates errors, and if all goes to plan, we gain the mechanism to adjust weights and correct these errors. This process also goes from the back of the network toward the input layer. It actually sends back the output that has already passed through the activation functions and adjusts the weights (an increase or decrease) to reduce errors and close down the gap between desired output and model output. Here, the key feature of ANNRW is to modify the weight gradually, this is called “learning rule.” This goes again and again till we get the minimum error. Error is the difference between desired correct output and actual output that is found from each iteration.

Methods’ Particularities

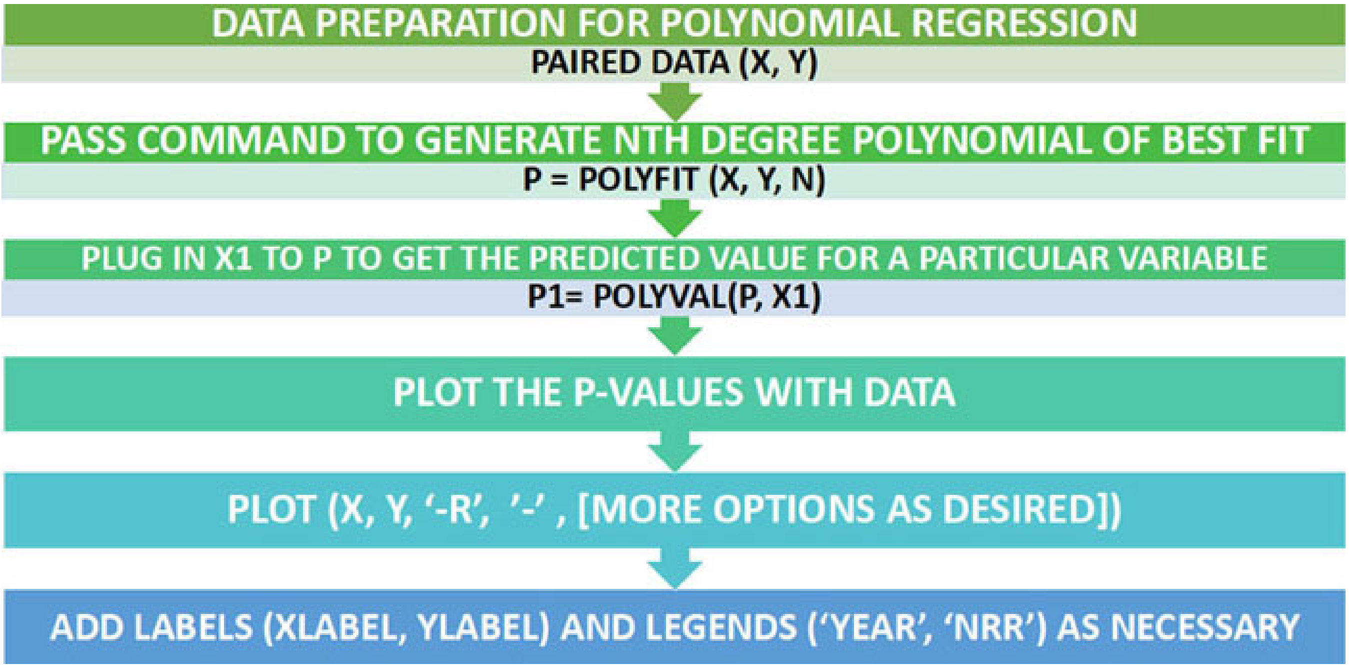
In practice, we must have the datasets for input (Xi), number of layers (Li), activation function (Sigma), learning rate (Alpha), iteration/epoch (I), and output (Yi). In our case, we have 47 parameters of interest and, so, we are using 15 hidden layers (as per rule, less than half of the parameters). The learning rate and activation function will be randomly chosen, and we will run the model 5,000 times with five outputs (levels of radon from 0–99, 100–199, 200–599, 600–1,199 and ≥1,200 Bq/m^3).

MPNR in Detail

For the polynomial regression, we used polyfit, polyval, and plot functions at curve fitting tool (cftool) in the same MATLAB (Figure 3). Initially, the vector of coefficients of x and y were developed from model fit equations. For example, for our input variable x, part of the vector was $v = [3\ -2\ 0\ 1\ -6\ 5]$ that is equivalent to a mathematical equation, $p = 3x^5 - 2x^4 + x^2 - 6x + 5$ (here the x^3 is absent, as is represented by 0 in the vector).

Similarly, the outcome variable (y) was NRR (not rounded radon level). Thus, our commands for polynomial were polyfit (x, y, n), where n denotes the degree of polynomial that best fitted the data.

Figure 3. Multiple/polynomial regression in MATLAB.



Source: Authors.

We used polyval command to plug any number that represented a particular year, for example, 2017 (x1) into the polynomial model as necessary, for example, p1 = polyval (p, x1) to get the prediction for that particular year. Then, we used the plot command to visualize the data such as plot ("Year," "NRR," [plus plot character options as desired]). Then, we added labels for the variables, for example, xlabel ("Year") and ylabel ("NRR"); legend ("Not Rounded Radon Prediction," 'Prediction per year'). In our model, ANNRW provided more precise and robust outputs compared to MPNR.

Section Summary

- ANNs are an emulation of the function of human neurons, with some fundamental differences.
- Machine learning ANN has evolved over time, where the new-generation ANNRW proved to be more powerful than the traditional ANN.
- We can run polynomial regression in MATLAB to compare the prediction outputs from ANNRW.

Practical Lessons Learned

IAQ is influenced by a large number of complex variables. While individual variables, once measured quantitatively, can be correlated with specific IAQ contaminant levels (in our case, radon levels), integration is challenging across highly diverse geographic, behavioral, and demographic areas. We had already performed a multivariate analysis of property metrics that significantly influence radon levels in our survey region (Stanley et al., 2019), finding a subset of 5–8 metrics that could predict (to a certain extent) whether a residential building might contain >100 Bq/m^3 or >500 Bq/m^3. However, this analysis, although useful, was not satisfactory enough in terms of knowing exactly what build features actually cause radon levels within the North American housing stock to increase as a function of build year. From a practical perspective, we knew a substantially more sophisticated approach is required. That is why we are now exploring the utility of a complex multilayer/deep neural network algorithm of AI to solve these data pattern recognition issues, building on the experiences of other researchers in the field (Kulikajevas et al., 2019; Siuly & Zhang, 2020).

While still underway, we have already learned several lessons. The new-generation ANNRW has proved to be more powerful than the traditional ANN. Another observation is that ANN uses the function almost similar to human brain to come to the most precise outcome. For example, when an issue is complex and one input depends on the outcome of other(s), we take more time to think and analyze the matter. Similarly, when there are interactions between some variables, more hidden layers need to be added. A simple example can clarify this: in the case of image recognition, at first ANN determines the patterns of local contrast (pixels); then, the particular features of an organ (nose or eye) are determined, after which the whole face is recognized. This is how it can confidently come to the accurate recognition of an individual and it is the strength of this application. We are applying similar approaches here, where the first ANN determines patterns of exposure by singular metric of basement type, then the particular features of a given building construction practices, then progressing on to the whole features of housing stocks, and to do so as a function of the evolving regional build codes, human behaviors, geologies, meteorology, and so on. It will be tremendously exciting as this approach begins to deliver concrete answers. This is the strength of this method but one weakness is that one has to clean and prepare the data, finding the regression coefficients for setting vectors of parameters, work meticulously on all the stages of setting and modifying the weights, activation functions, learning rates, identify and quantify any interactions between them in a very efficient manner that can achieve after hundreds of trials training and cross-validations before coming up with a useful model that performs with the maximum precision.

Section Summary

- ANN can be applied in research where precision is crucial in solving a real-life problem requiring data pattern recognition, and not possible using traditional approaches.
- Getting precise ANN outcome is not a straightforward task, it needs rigorous data preparation, mathematical and statistical shrewdness, and computer programming along with a clear understanding of the rules of artificial intelligence.

Conclusion

Maintaining IAQ is crucial to maintain the good health of people living in highly developed cold countries where populations spend the majority of their lives within indoor built environments (Klepeis et al., 2001; Setton et al., 2013). Radon exposure, the second leading cause of all lung cancers and the leading cause in nonsmokers, is a substantial public health problem and is a highly attractive target for cancer prevention strategy development (World Health Organization, 2009). Given the complexity of radon exposure variables (Stanley et al., 2019), it is now essential to apply the most sophisticated analytic strategies to comprehend the nature and factors of radon exposure within the built environment, to actually engineer it out of properties in the near future. In this case study, we tested and compared a new-generation machine learning and separate statistical models as a function of multiple radon predictors. Error margins were compared between two models for the first time using ANNRW and MPNR. We proved that ANNRW is a reliable and efficient method that can precisely determine radon risk level. This is contributing to a more informed, evidence-based set of policy recommendations with the potential to prevent the most lethal type of lung cancer and save thousands of potential lives and enormous costs to health care system.

Section Summary

- IAQ is crucial for the health of a population.
- Designing interventions to reduce a health risk requires precise estimation of the risk.
- Application of new-generation ANN of machine learning algorithm can successfully reduce error in indoor radon risk prediction.

Classroom Discussion Questions

Classroom Discussion Questions

1. Why research with indoor air quality is important for the health of a population?
2. What are the possible biases in selecting the participants for research project?
3. What is the usefulness of applying machine learning algorithms in research in general and radon exposure in particular?
4. What combination of knowledge base is required to be able to successfully apply machine learning algorithm in health research?

Funding Statement

SMK was supported by the University of Calgary Eyes High Post-Doctoral Scholar program. AAG is the Canada Research Chair for Radiation Exposure Disease and this work was undertaken, in part, thanks to funding from the Canada Research Chairs program. The AAG laboratory is funded by CIHR and NSERC and this work was made possible by the Alberta Real Estate Foundation.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

Further Reading

Van-Gerven, M., & **Bohte, S.** (Eds.) (n.d.). Artificial neural networks as models of neural information processing. https://www.frontiersin.org/research-topics/4817/artificial-neural-networks-as-models-of-neural-information-processing#articles

Van-Gerven, M., & **Bohte, S.** (2017). Editorial: Artificial neural networks as models of neural information processing. *Frontier in Computational Neuroscience*, 11, Article 114. https://doi.org/10.3389/fncom.2017.00114

Le, T. T., **Pham, B. T.**, **Ly, H. B.**, **Shirzadi, A.**, & **Le, L. M.** (2020). Development of 48-hour precipitation forecasting model using nonlinear autoregressive neural network. In **C. Ha-Minh**, **D. Dao**, **F. Benboudjema**, **S. Derrible**, **D. Huynh**, & **A. Tang** (Eds.), *CIGOS 2019, Innovation for sustainable infrastructure* (Lecture Notes in Civil Engineering, Vol. 54, pp. 1191–1196). Springer. https://doi.org/10.1007/978-981-15-0802-8_191

Hayakawa, Y., **Oonuma, T.**, **Kobayashi, H.**, **Takahashi, A.**, **Chiba, S.**, & **Fujiki, N. M.** (2020). *Feature extraction of video using artificial neural network*. IGI Global. https://www.igi-global.com/chapter/feature-extraction-of-video-using-artificial-neural-network/237905

Macdonald, C. (2017). Machine learning fundamentals (I): Cost functions and gradient descent. *Toward Data Science* [Blog]. https://towardsdatascience.com/machine-learning-fundamentals-via-linear-regression-41a5d11f5220

References

Barnett, P. J., & **Bajc, A.** (2002). Quaternary geology. In **D. H. Rousell** & **K. J. Jansons** (Eds.), *The physical environment of the City of Greater Sudbury* (pp. 56–85). Ontario Geological Survey.

Darby, S., **Hill, D.**, **Auvinen, A.**, **Barros-Dios, J.**, **Baysson, H.**, **Bochicchio, F.**, … **Doll, R.** (2005). Radon in homes and risk of lung cancer: Collaborative analysis of individual data from 13 European case-control studies. *British Medical Journal*, 330(7485), 223–230. https://doi.org/10.1136/bmj.38308.477650.63

Garrity, C. P., & **Soller, D. R.** (2009). *Database of the geologic map of North America: Adapted from the map by J.C. Reed, Jr., & others (2005): U.S. geological survey data series 424.* http://purl.access.gpo.gov/GPO/LPS116275

Gaskin, J., **Coyle, D.**, **Whyte, J.**, & **Krewksi, D.** (2018). Global estimate of lung cancer: Mortality attributable to residential radon. *Environmental Health Perspective*, 126(5), Article 057009. https://doi.org/10.1289/EHP2503

Goodarzi, A. A., **Anikin, A.**, & **Pearson, D. D.** (2016). Environmental sources of ionizing radiation and their health consequences. In **I. Kovalchuk** & **O. Kovalchuk** (Eds.), *Genome stability* (pp. 569–581). Elsevier.

Grundy, A., **Brand, K.**, **Khandwala, F.**, **Poirier, A.**, **Tamminen, S.**, **Friedenreich, C. M.**, & **Brenner, D. R.** (2017). Lung cancer incidence attributable to residential radon exposure in Alberta in 2012. *CMAJ Open*, 5(2), E529–E534. https://doi.org/10.9778/cmajo.20160053

Health Canada. (2012). *Cross-Canada Survey of Radon Concentrations in Homes—Final report*. Government of Canada.

Klepeis, N. E., **Nelson, W. C.**, **Ott, W. R.**, **Robinson, J. P.**, **Tsang, A. M.**, **Switzer, P.**, … **Engelmann, W. H.** (2001). The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *Journal of Exposure Analysis and Environmental Epidemiology*, 11(3), 231–252. https://doi.org/10.1038/sj.jea.7500165

Krewski, D., **Lubin, J. H.**, **Zielinski, J. M.**, **Alavanja, M.**, **Catalan, V. S.**, **Field, R. W.**, … **Wilcox, H. B.** (2005). Residential radon and risk of lung cancer: A combined analysis of 7 North American case-control studies. *Epidemiology*, 16(2), 137–145. https://doi.org/10.1097/01.ede.0000152522.80261.e3

Kulikajevas, A., **Maskeliūnas, R.**, **Damaševičius, R.**, & **Misra, S.** (2019). Reconstruction of 3D object shape using hybrid modular neural network architecture trained on 3D models from shape net core dataset. *Sensors*, 19(7), Article 1553. https://doi.org/10.3390/s19071553

Lee, J.-G., **Jun, S.**, **Cho, Y.-W.**, **Lee, H.**, **Kim, G. B.**, **Seo, J. B.**, & **Kim, N.** (2017). Deep learning in medical imaging: General overview. *Korean Journal of Radiology*, 18(4), 570–584.

Moore, S., **Stanley, F. K. T.**, & **Goodarzi, A. A.** (2014). The repair of environmentally relevant DNA double strand breaks caused by high linear energy transfer irradiation—No simple task. *DNA Repair*, 17, 64–73. https://doi.org/10.1016/j.dnarep.2014.01.014

Oberle, K. M., **Page, S. A.**, **Stanley, F. K. T.**, & **Goodarzi, A. A.** (2019). A reflection on research ethics and citizen science. *Research Ethics*, 15, 1–10.

Papastefanou, C., **Stoulos, S.**, **Manolopoulou, M.**, **Ioannidou, A.**, & **Charalambous, S.** (1994). Indoor radon concentrations in Greek apartment dwellings. *Health Physics*, 66(3), 270–273.

Setton, E., **Hystad, P.**, **Poplawski, K.**, **Cheasley, R.**, **Cervantes-Larios, A.**, **Keller, C. P.**, & **Demers, P. A.** (2013). Risk-based indicators of Canadians' exposures to environmental carcinogens. *Environmental Health*, 12, Article 15. https://doi.org/10.1186/1476-069X-12-15

Siuly, S., & **Zhang, X.** (2020). Guest Editorial: Special issue on application of artificial intelligence in health research. *Health Information Science and Systems*, 8(1), Article 1. https://doi.org/10.1007/s13755-019-0089-x

Stanley, F. K. T., **Irvine, J. L.**, **Jacques, W. R.**, **Salgia, S. R.**, **Innes, D. G.**, **Winquist, B. D.**, … **Goodarzi, A. A.** (2019). Radon exposure is rising steadily within the modern North American residential environment and is increasingly uniform across seasons. *Scientific Reports*, 9, Article 18472.

Stanley, F. K. T., **Zarezadeh, S.**, **Dumais, C. D.**, **Dumais, K.**, **MacQueen, R.**, **Clement, F.**, & **Goodarzi, A. A.** (2017). Comprehensive survey of household radon gas levels and risk factors in southern Alberta. *CMAJ Open*, 5(4), E255–E264. https://doi.org/10.9778/cmajo.20160142

Wilson, D. L., **Gammage, R. B.**, **Dudney, C. S.**, & **Saultz, R. J.** (1991). Summertime elevation of ^{222}Rn levels in Huntsville, Alabama. *Health Physics*, 60(2), 189–197.

World Health Organization. (2009). *WHO handbook on indoor radon: A public health perspective*.