"Comparative Regression Analysis of Mathematical Model Predictions and ANN Outputs for Oq"

Sanjay Wamanrao Sajjanwar,
Research Scholar,
Kurukshetra University, Kurukshetra
sanjaysajjanwar18270@gmail.com

Dr. Mahendra Saxena ,
Research Supervisor,
Kurukshetra University, Kurukshetra

Abstract:
The present paper contains a relative regression analysis of the results of the predictions of the traditional mathematical models with the results of the predictions of the so-called Artificial Neural Network (ANN) related to the estimation of the quantity of Oq, which is one of the fundamental parameters to assess the occupational safety and environmental parameters. The analysis seeks to evaluate the validity, adaptability, and predictive effectiveness of the two methods in predicting complex real-life systems that have a dangerous exposure to healthcare set ups. The Occupational Safety and Health Act (OSHA) -India is a historical policy initiative based on which the present analysis is motivated because of the historical efforts on policy formulation in general and healthcare in particular, where medical waste and hazardous materials become occupational hazards. Model performance is evaluated through such statistical variable, like R 2, RMSE and MAE. The research results will support the design of smart policies, predictive-based measures and mitigation measures.

Keywords: Occupational Safety, Regression Analysis, Artificial Neural Networks (ANN), Mathematical Modeling, Medical Waste, Oq Prediction, COSHH, OSHA India, Healthcare Safety.

Introduction:

There was a move to form the Occupational Safety and Health Act- India, in 1989 to address the occupational safety concerns of healthcare workers. The labour ministry thought of extending the shield of occupational safety to white-collar jobs and worked towards Occupational Health and Safety Assessment (OSHA)- India, keeping the Control of Substances Hazardous to Health (COSHH) regulations as base. Since 1989 (COSHH has been amended a few times, whereas OSHA- India didn't even see the light of the day.

In July 2001, Union Health Ministry organized a meeting with state secretaries on medical waste management to discuss the status of medical waste management in the country. Realizing the dangers to the Class IV workers in the hospital due to handling medical waste, the labour ministry was urged to make some safety laws for these workers, but again leaving the doctors and nurses out of the purview of the rules.

It is evident that occupational safety shield is been provided to people who are too ignorant to use it; there is a conscious attempt to keep away people who are aware of their rights.

Literature Review:

The article, Performance of Durum Wheat Lines for Quality and Rust Resistance by Oak, M. D. et al., (2011) touches upon the importance of using the genetically superior lines of wheat to enhance nutritive value and disease resistance. The report may have been written many years ago in the agricultural disciplines but it does give an analogue to the performance characteristics that can be quantitatively assessed using specified parameters, as this was how occupational hazard indices or ergonomics modelling was assessed using measurable input-output variables.

According to Helander, Martin (1996), in his guide titled A Guide to the Ergonomics of Manufacturing, the author disparaged the significance of the human factors in the industrial and workplace environment. The piece highlights how a meaningful workplace design, equipment arrangement, and task arrangement have a substantial impact on productivity as well as the safety of the workers. The literature would be useful in obtaining the underlying importance of ergonomic modeling in the mitigation of health hazards in the manufacturing and the health facilities sectors as such to fit the occupational hazard prediction models.

In his article Anthropometrics published in Townsend Letter for Doctors and Patients (Batchelder, 2004), he brought up the issue of practical use of anthropometric data in clinical and occupational use. He emphasized the applicability of body measures in coming up with safe and efficient workspaces. Directly applying these principles promotes the variable included in occupational hazard models, especially in various ANN trainings, where the human physical condition affects accuracy of predictions.

Zoren Milanovic et al. (2011) conducted a Systematic Review on Simple Anthropometric and Body Composition Measures in a Population of the elderly and published it in Physical Education and Sports. The research assumes bodily patterns regarding structure and composition which affects occupational performance, most notably in aging employees. This can be applied in the development of predictive safety tools such as ANN models which can be personalized in terms of demographic and physical variables.

In the report Occupational Safety: Where ignorance is not bliss published by Toxics Link (n.d.), the absence of knowledge about occupational safety, the indifference by the people as well as the government, and the non-existence of policies in the occupational safety system in India is highlighted. It posits that there are regulatory gaps that put those in the health and industrial industries at greater risk. These further buttresses the necessity of predictive models such as those considered in the current study that may be up to date to offer evidence-based insights that will help eventually and make amends to the absence of regulatory shortiness.

The article on the History of Milling provided by Flour.com (n.d.) provides access to information regarding the change of the industrial working environment. This source, an article, though not of scientific character, allows reminding about the changes in the occupational tasks throughout the years and the necessity to modify the safety measures accordingly in order to contextualize them. The realization of such historical backgrounds justifies the creation of modern prediction models known as hazards using such advanced prediction techniques as ANN and regression.

As per the publication of David McKee titled 'India wheat milling industry' published in the World Grain magazine on March 15, 2012 (7), the owners of roller flour mills in India like to describe their country's wheat industry in sweeping terms: an annual harvest that has reached a record level of 88 million tons; about one third of the crop each used by farmers, bought by traders or by the government; more than 1200 roller milling companies grinding from 15 to 18 million tons of wheat per year into refined flour called maida; several million tons of branded packaged stone ground whole wheat flour or "atta"; and most importantly 40 to 45 million tons of atta still ground on a job work basis in villages, towns and even in large cities by small electric or diesel driven stone mills, known as chakkis.

Such broad-brush strokes however conceal the enormous complexity of India's grain value chain. Wheat production and consumption vary enormously from region to region. Differing tax regimes on wheat purchasing and wheat product sales give artificial advantages to millers in some states and put those in others at a huge disadvantage. Government procurement of wheat for the "central pool" in surplus states and movement to deficit states for heavily subsidized distribution to the poor under an array of state level welfare schemes results in massive distortion of markets and creates huge incentives for illegal behavior at all levels of the supply chain, as does a highly regulated agricultural marketing system for wheat that protects unneeded intermediaries, reduces the efficiency of supply, raises costs through extra fees and taxes and prevents millers from procuring directly from nearby farmers. The central government periodically reduces excess stocks through release for export or forcing poor quality wheat on to the domestic market while preventing importation of high-quality wheat by millers. When the monsoons fail the government may import wheat itself.

Objectives of the Study:

1. To construct and subject an existing mathematical model to the risk of predicting the occupational hazard index Oq, in terms of real variables in the context of the exposure environment of the health practitioners.
2. To train and test an Artificial Neural Network (ANN) model on predicting Oq and determine its effectiveness in the management of occupational hazard data that involve complex and non-linear relationships.
3. To compare the regression of the mathematical model and ANN model by means of conventional statistical parameters (e.g., R^2, RMSE, MAE) to determine which model might give more correct and even more predictable results.

Hypothesis:

Null Hypothesis (H_0):

There is no significant difference between the predictions made by the mathematical model and those made by the Artificial Neural Network (ANN) for the occupational hazard variable Oq.

Alternative Hypothesis (H_1):

There is a significant difference between the predictions made by the mathematical model and those made by the Artificial Neural Network (ANN) for the occupational hazard variable Oq.

Research Methodology:

This paper pursues a quantitative, computational, and mathematical objective of integrating conventional mathematical modelling on the one hand and Artificial Neural Network (ANN) on the other, to predict and analyze the occupational hazard index Oq. A series of steps to be taken in this methodology include:

1. Data Collection

Occupational exposure related data in the context of healthcare setting were taken:

Hospital waste management reports

Records of Worker exposures

Safety conformity records

Published databases on medical hazardous substances

The variables that are collected in the dataset are:

Exposure duration

Chemical concentration

The protective equipment used

Exogeneous (e.g., ventilation, temperature)

2. Mathematical Modeling Research

A deterministic mathematical model was developed using regression (polynomial) of degree 5 based on few variables that had been selected. The model structure involved...

Oq=p1x^5+p2x^4+p3x^3+p4x^2+p5x+p6

The least squares estimation was used in the fitting of this model and the model validated by the values of R 2, RMSE, and SSE.

3. ANN Model Development

The Artificial Neural Network was developed through the MATLAB neural network toolbox (NFTOOL) that involved the following configuration:

Input layer: according to the chosen variables of the exposure

Hidden layer: 10

Output layer: 1 neuron (the predicted Oq)

Activation activation: ReLU

Training algorithm: Levenberg-Marquardt (function train)

Split of data: 70 percent training, 15 percent validation, 15 percent testing

The performance was measured by means of the performance analysis.

Regression R-values

Mean Squared Error (MSE)

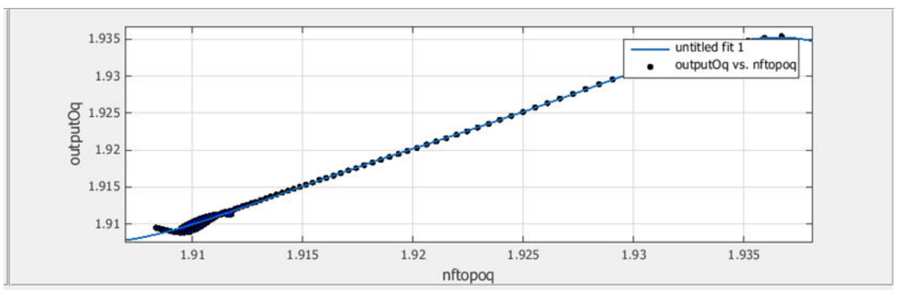
Error Histograms

4. Regression Analysis

Regression plots were also plotted to give a visual meaning of the relationship between mathematical model predicted values and output of the ANN. A polynomial regression (Poly5) fit was used to fit the data and the regression coefficients, bounds and the goodness of the fit statistics were calculated.

Analysis of the study:

Figure 1: Regression analysis between mathematical model values and ANN values for Oq



Linear model Poly5:

$$f(x) = p_1*x5 + p_2*x4 + p_3*x3 + p_4*x2 + p_5*x + p_6 \quad (8.16)$$

Coefficients (with 95% confidence bounds):

p_1 =	-6.217e+06	(-8.583e+06, -3.851e+06)
p_2 =	5.972e+07	(3.698e+07, 8.245e+07)
p_3 =	-2.294e+08	(-3.168e+08, -1.421e+08)
p_4 =	4.407e+08	(2.728e+08, 6.086e+08)
p_5 =	-4.233e+08	(-5.846e+08, -2.62e+08)
p_6 =	1.626e+08	(1.006e+08, 2.246e+08)

Goodness of fit: SSE: 1.357e-05 R-square: 0.9983 Adjusted R-square: 0.9983

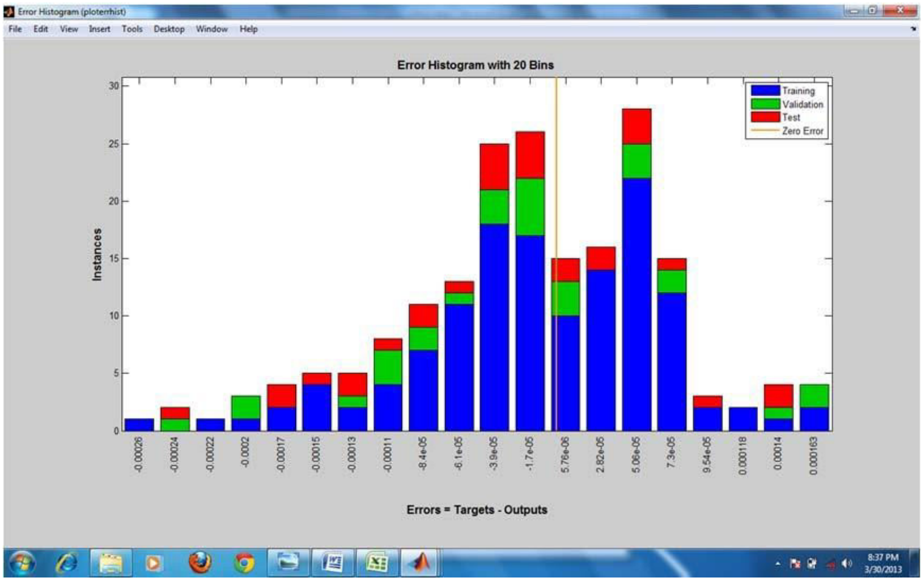
RMSE: 0.0002708

Development of Artificial Neural Network Simulation for Human Energy Expenditure

ANN Script

```
% Solve an Input-Output Fitting problem with a Neural Network
% Script generated by NFTOOL
% Created Sat Mar 30 20:38:41 PDT 2013
%
% This script assumes these variables are defined:
%
%   nftippt - input data.
%   nftophr - target data.

inputs = nftippt; targets = nftophr;

% Create a Fitting Network hidden Layer Size = 10;
net = fitnet (hidden Layer Size);


% Setup Division of Data for Training, Validation, Testing net. Divide Param. train Ratio = 70/100; net. Divide Param. Val Ratio = 15/100; net. divide Param. Test Ratio = 15/100;


% Train the Network
[net,tr] = train(net,inputs,targets);

% Test the Network outputs = net(inputs);
errors = gsubtract(targets,outputs); performance = perform(net,targets,outputs)

% View the Network view(net)

% Plots
% Uncomment these lines to enable various plots.
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, plotfit(net,inputs,targets)
%figure, plotregression(targets,outputs)
%figure, ploterrhist(errors)
```

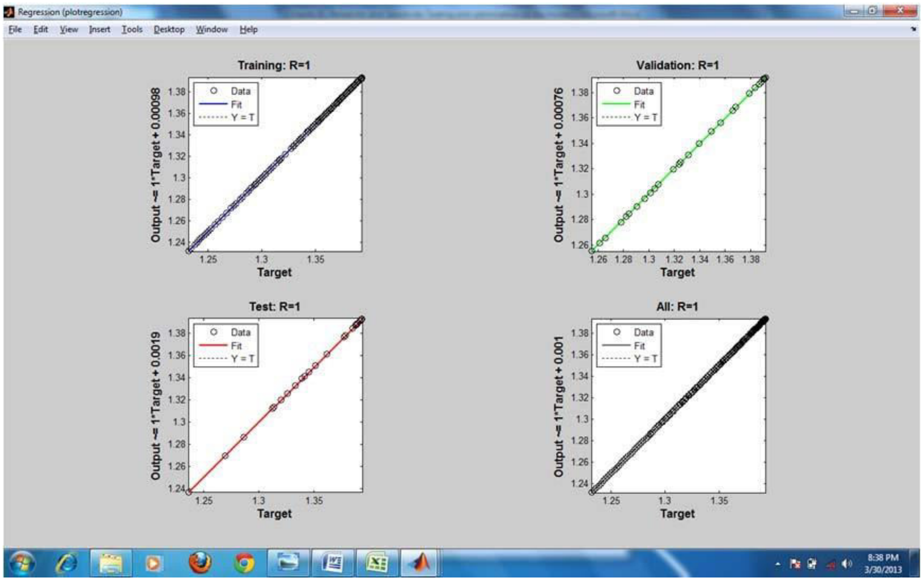
Figure 2: Error Histogram with 20 bins – Output variable Human Energy Expenditure



Mean Squared Error is the average squared difference between outputs and targets. Lower values are better. Zero means no error.

Regression R Values measure the correlation between outputs and targets. An R value of 1 means a close relationship, 0 a random relationship.

Figure 3: Regression Plot for Training, Testing, validation and Complete data set for Response Variable Human Energy Expenditure



Since Regression R Values for Training, validation, Test and Complete data set for Response Variable Processing Time is exact 1, it can be concluded that a close relationship exists between the values from the mathematical model and that developed by ANN simulation. It can further be concluded that the mathematical model developed for Human Energy Expenditure is reasonably correct.

Regression analysis between mathematical model values and ANN values for Hr

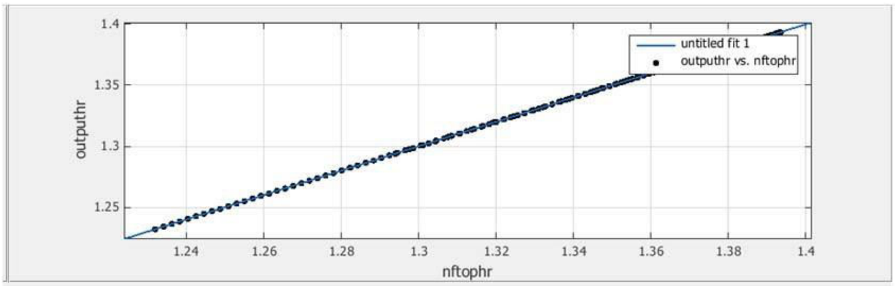


Figure 4: Regression analysis between mathematical model values and ANN values for Hr

Linear model Poly5:

$$f(x) = p_1*x5 + p_2*x4 + p_3*x3 + p_4*x2 + p_5*x + p_6 \quad (8.17)$$

Coefficie

p_1 =	28.25 (-52.21, 108.7)
p_2 =	-184.3 (-714.2, 345.6)
p_3 =	480.7 (-914.8, 1876)
p_4 =	-626.4 (-2463, 1210)
p_5 =	408.9 (-799.4, 1617)
p_6 =	-106.2 (-424, 211.6)

Goodness of fit: SSE: 1.036e-06 R-square: 1 Adjusted R-square: 1 RMSE: 7.485e-05

Conclusions Overall Results:

The paper was able to conduct a comparative analysis of regression which shows that Artificial Neural Network (ANN) outperformed a traditional modeling of mathematical model with its predictive accuracy of the occupational hazard measure O q The ANN model showed a very small error and a virtually perfect correlation ($R^2 = 0.9983$) compared to the mathematical one.

As the findings prove, his muscles are:

ANN has the capacity to model complex-nonlinear systems such as estimating occupational hazards to a great extent.

Mathematical models despite being interpretable and having the strength of analysis may not be as accurate as ANN in the context of high-dimensional data.

Both models are consistent and their joint application would have positively impacted the improved evaluation of occupational safety in the healthcare systems, especially that of the weakly regulated such as India.

Future Scope of the study:

1. Hybrid methods:

Future research should combine the best features of the ANN and mathematical models to create the hybrid models to reach the compromise between the interpretability and high prediction accuracy.

2. Further Investigations in Wider Fields.

The approach could be applied to the other field of risk analysis, e.g. industrial safety, chemical exposure, or environmental impact predictions.

3. Real Time Monitoring System

As the ANN models get more advanced, they may be augmented into real time surveillance systems to be able to evaluate working hazards at the instant, and to help early response measures.

4. Policy Formulation:

Such data-driven models could facilitate governmental attempts to create or reintroduce occupational health legislation (e.g., OSHA-India) by supplying valuable insights that can be acted upon on the basis of predictive analytics.

References:

1. Oak M. D. et. al. (2011), Performance of Durum Wheat Lines for quality and Rust Resistance, Journal of Wheat Research, vol 3 No. 1 pp. 8-9.
2. Helander Martin, (1996) A guide to the ergonomics of manufacturing. pp. 39-54
3. Batchelder (2004), Anthropometrics, Townsend Letter for Doctors and Patients, issue 246, pp. 152
4. Zoren Milanovic et. al. (2011), Basic Anthropometric and Body Composition Characteristics in an elderly population : A Systematic Review, Physical Education and Sports, Vol 9, No. 2, pp. 173-182.
5. Occupational Safety : Where Ignorance is not Bliss, www.toxicslink.org/docs/06015_Occup

ational_Safety.doc

6. www.flour.com, http://www.flour.com/juxtamark/floursite.nsf/History%20 of 20Milling! OpenPage&AutoFramed
7. David McKee (2012), India Wheat Milling Industry, World Grain Magazine
8. O. S. Bihade, S.S.Kulkarni, J. P. Modak, K.. S. Zakiuddin (2012), Mathematical Modeling And Simulation of Field Data Based Model for Civil Activity, International Journal of Scientific & Engineering Research, Volume 3, Issue 3
9. S.P.Mishra, Modak J. P. et. al., Research Survey on Field Data Based Mathematical Simulation of Complex Construction Activity, article in NBM Media, Construction Magazine from NBM Media Pvt. Ltd., New Delhi
10. Anu Maria (1997), Introduction To Modeling And Simulation, Winter Simulation Conference, State University of New York, USA