Applied Science and Innovative Research
ISSN 2474-4972 (Print) ISSN 2474-4980 (Online)
Vol. 9, No. 3, 2025
www.scholink.org/ojs/index.php/asir

Original Paper

Neural Network Implementation for CRC Awareness Prediction

Ale J. Hejase[1], Ali A. Haidous[2], Ahmad A. HejaseBazzi[3] & Hussin J. Hejase[4]

[1] Senior Researcher, Former Professor, Adnan Kassar School of Business (AKSOB), Lebanese American University, Beirut, Lebanon

[2] Adjunct Professor, Edge AI Department, University of South Alabama, Mobile, AL, USA

[3] Department of Computer Science & Engineering, Michigan State University, East Lansing, MI, USA

[4] Basic and Applied Sciences Research Center, Al Maaref University, Beirut, Lebanon

Corresponding author: Hussin J. Hejase, E-mail: hussin.hejase@mu.edu.lb

Abstract

Artificial Intelligence (AI) is prevalent, driven by the resurgence of machine learning (ML) and advancements in data-driven applications, which have created new opportunities to surpass traditional statistical limits in health analytics. This study develops and assesses a deep artificial neural network (ANN) to predict public awareness of Colorectal Cancer (CRC) using six demographic and socioeconomic factors: age, gender, marital status, educational level, work status, and place of residence. Primary data were collected through a survey administered to a sample of 1,229 Lebanese individuals.

We systematically describe the data preprocessing steps, including handling missing values, encoding categorical variables, and scaling features. We specify the model architecture, which includes input, output, and one or more hidden layers, along with training optimization methods such as class weighting, early stopping, and learning rate scheduling. The model is evaluated using thorough cross-validation and held-out testing, with metrics like discrimination and calibration assessments, including accuracy percentages and confidence scores. An explainability technique, including feature importance analysis, is employed to enhance transparency. The proposed network demonstrates reliable binary classification performance while emphasizing interpretability, deployment readiness, and the reproducible integration of machine learning into healthcare analytics.

Keywords

Awareness, CRC, ANN, ML, Demographics, Lebanon

1. Introduction

Colorectal cancer (CRC) remains a leading cause of cancer deaths worldwide (Rawla, Sunkara, & Barsouk, 2019), emphasizing the urgent need for accessible and affordable screening options (A. Hejase, Hejase, Kassem, & Hejasebazzi, 2025). This relies heavily on well-crafted awareness campaigns that actively engage and educate the public (Van Altena, 2021). Successful CRC awareness efforts focus on promoting screening, raising awareness about risk factors and symptoms, and utilizing various communication channels (Nemer et al., 2016; Ranganath, Lujaina, Al Saidi, et al., 2022). CRC is a key contributor to global cancer mortality (Li, Geng, Luo, Wang, et al., 2024), highlighting the urgent need for optimized and cost-effective screening methods.

Colorectal Cancer (CRC) awareness varies widely across different demographic groups (Pankratz, Kanda, Kosich, Edwardson, et al., 2024). Understanding what influences CRC awareness is essential for creating targeted health education campaigns (Eka Rusdi Antara, 2024). Getting screened regularly, starting at age 45, is key to preventing colorectal cancer (cancer of the colon or rectum). According to Abreu Lopez et al. (2024), the American Cancer Society (2024), and A. Hejase et al. (2025), "Individuals aged 45 to 75 should undergo regular screenings for colorectal cancer. People under 45 who believe they might be at high risk for colorectal cancer, along with those over 75, should discuss screening options with their doctor." CRC often shows no symptoms, especially in its early stages. That's why regular screenings are so important. Cleveland Clinic (2022) warns that sometimes, "unusual growths known as polyps develop in the colon or rectum. Eventually, some polyps can turn into cancer. Screening tests can detect polyps, allowing for their removal before they become cancerous." Screening helps find colorectal cancer early, when treatment is most effective (CDC, 2025).

1.2 Problem Statement and Objectives

Over the past decade, the resurgence of machine learning (ML) and advancements in artificial intelligence (AI), driven by data science, present unique opportunities to overcome traditional statistical and research limitations, improving health data analytics to address healthcare challenges better and ultimately enhance health service delivery (Hassibi, 2016; Stewart, 2019). This paper offers a comprehensive implementation of a deep neural network for predicting CRC awareness based on demographic and socioeconomic factors. The proposed approach includes data preprocessing, model architecture design, and training optimization. The neural network is expected to perform robustly in binary classification tasks, employing appropriate regularization techniques to prevent overfitting and ensure good generalization. Its goal is to solve the binary classification problem of predicting whether an individual has heard of CRC based on six demographic features. The primary objectives of this implementation are:

1) Develop a robust neural network classifier for CRC awareness prediction.
2) Implement proper data preprocessing and feature scaling techniques.
3) Create a production-ready API (Application Programming Interface) for real-time predictions.
4) Provide comprehensive model analysis and visualization tools.
5) Ensure a scalable deployment architecture.

1.3 Merit of the Research

This paper has several notable strengths. First, it contributes original data from Lebanon, a region largely underrepresented in the existing literature on this specific subject. To our knowledge, no prior studies have addressed this topic within the Lebanese context, making this research both timely and innovative.

Additionally, the sample size is sufficiently large to lend credibility and facilitate broader generalization of the findings. Overall, it's a meaningful addition to the academic conversation.

2. Materials and Methods

This paper uses a quantitative, exploratory, and explanatory approach. With a positivist stance and survey strategy, primary data is collected and analyzed.

2.1 Sampling and Sample Size

The authors utilized non-probabilistic convenience sampling in their study. Respondents were chosen based on their willingness to participate after being informed that they could withdraw at any time without question asked. Participants have free will. They are told that their responses are confidential, with no personal information revealed, and their feedback is used for academic research. Cochran's formula was applied to determine the needed number of participants (Hejase & Hejase, 2013). With an approximate 4% margin of error, a 99% confidence level, and a 50% attribute distribution (i.e., gender), the minimum sample size was estimated at 1,112 participants. Furthermore, for the sake of validation, a reliability error value was determined utilizing a method similar to that described in the works of various researchers (Rammal et al., 2024, 2025; A. Hejase et al., 2025; Chehimi and Hejase, 2024; Masoudi and Hejase, 2023; H. Hejase et al., 2023a, b). For a population of 5,000 or more (at $\alpha = 5\%$) and a sample size over 1,000, the margin of error for this study at the 95% confidence level is roughly ± 3.5%. That indicates that in 96.5 out of every 100 times the survey is repeated, the outcomes will differ by at most 3.5%. This degree of dependability is perfect for this investigative study.

2.2 Questionnaire Design

This study questionnaire used the same form as A. Hejase et al. (2025). It is divided into two sections, whereby section 1 delineates awareness of CRC indicators, symptoms, risk elements, and readiness for early screening initiatives (focusing primarily on colonoscopy and FIT tests). Section two constitutes the socio-demographic traits of participants (Age, gender, education level, location of residence, work status, and marital status). The questionnaire was distributed through convenient social connections between February 2025 and May 2025. The paper questionnaire was administered to a sample of 1,250 individuals; however, 1229 valid questionnaires resulted. Therefore, a response rate of 98.32% was obtained.

2.3 Data Analysis

A neural network (ANN) modeling calculations and considerations are based on the collected responses from the aforementioned questionnaire. An ANN is a mathematical structure created to emulate the operations of the brain's central nervous system (Hadoua, Belaadi, Ghernaoutb, & Mukalazid, 2025). It can recognize patterns and learn the ability to make forecasts by utilizing empirical models (Hejase & Hejase, 1997). Neural networks serve a wide range of purposes. To identify the best parameters for the prediction model, it is crucial to conduct experiments on these models (Turco et al., 2021). The quantity of cells in the hidden layer is essential in network design. The current study examined different quantities of neurons in the concealed layer. The ideal quantity varies based on the particulars of the problem at hand and the training method employed, ascertained through a series of ongoing experimentation (Altıntaş et al., 2019). Therefore, data processing for neural networks involves preparing raw data into a format that the network can effectively learn from. This includes cleaning, transforming, and scaling the data, as well as potentially reducing dimensionality or creating new features. These steps are crucial for optimal network performance and generalization (Peace, Chris, & Victor, 2024). The adopted dataset contains the demographic information collected from the survey participants. As previously mentioned, the demographics set includes six input variables and one binary target variable as follows:

1. Age: Participant age at last birthday (Covariate, continuous variable).
2. Gender: Gender classification (Categorical: 1 = Female, 2 = Male).
3. Marital Status: Marital status (Categorical: 1 = Single, 2 = Married, 3 = Divorced, 4 = Separated, 5 = Widowed).
4. Education: Education level (Categorical: 1 = Higher Studies, 2 = University, 3 = High School, 4 = Less than High School)
5. Where You Live: Lebanese Geographic location (Categorical: 1 = Akkar, 2 = Mount Lebanon, 3 = Bekka, 4 = North Lebanon, 5 = Dahie, 6 = Beirut, 7 = South Lebanon, 8 = Shouf).
6. Work Status: Employment status (Categorical: 1 = Works, 2 = No-Work, 3 = Housewife, 4 = Retired, 5 = Studying).

One target variable is included

1. Heard of CRC: Binary outcome (Originally encoded as 1 = Yes, 2 = No)

2.3.1 Data Processing

For data preprocessing, the authors followed the steps recommended by ANN literature (IBM, 2019; Pisa, Santín, Lopez Vicario, Morell, & Vilanova, 2019). The key preprocessing steps were:

(1) Data Cleaning: Remove rows with missing values across all features. This step eliminated 10 records, leaving 1,219 valid data records.

(2) Target Encoding: Convert original encoding (1=Yes, 2=No) to binary (1=Yes, 0=No).

(3) Continuous Feature Scaling: Apply Standard Scaler (Standardizes age by removing the mean and dividing by the standard deviation of ages).

(4) Categorical Features: The method of one-hot encoding transforms categorical data into a numerical format suitable for use by machine learning (ML) algorithms. It operates by generating a binary column for every distinctive category in a feature, where a value of 1 signifies the existence of that category and 0 signifies its lack.

(5) Data Validation: Ensure all features are properly typed as integers.

(6) Data Partition: The 1,219 data cases were randomly partitioned into 81.5% (993) for training, 8.6% (105) for testing, and 9.9% (121) for validation.

2.3.2 Neural Network Architecture

Figure 1 illustrates that the architecture of a neural network is made up of an input, an output, and one or more hidden layers.

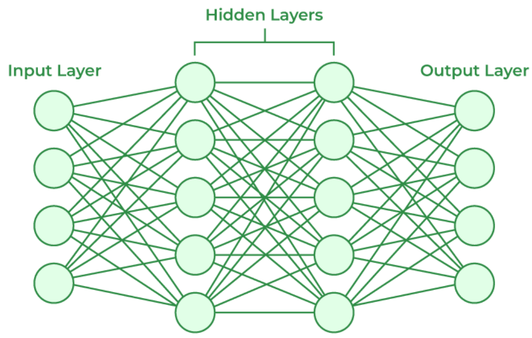

Figure 1. Neural Networks Architecture

Source: GeeksforGeeks, 2025a

Hadoua (2025) posits that "Artificial neural networks (ANNs) are a subset of machine learning designed to mimic the processing power of a human brain." ANNs operate by transmitting data through the levels of a synthetic neuron, as shown in Figure 2.

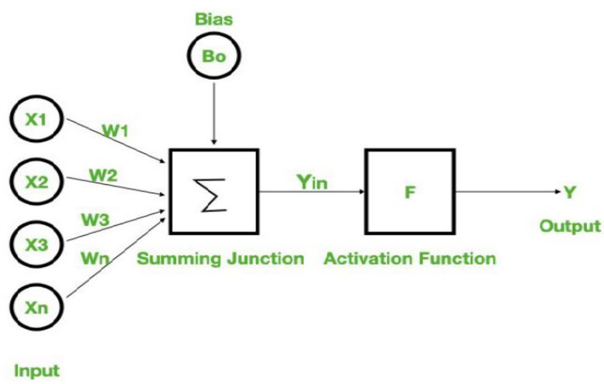

Figure 2. Artificial Neuron

Source: GeeksforGeeks, 2025b.

Every artificial neuron consists of these elements, as shown in Figure 2 and addressed in Exhibit 1.

Exhibit 1: Describing an artificial Neuron

(1) Input (X_i) refers to the data fed into the neuron for purposes of training and learning. (2) Weight (W_i) aids in structuring the input variables based on their significance and contribution impact. (3) A Summing Junction occurs when all the weighted inputs are aggregated and merged into a single intermediate output variable (Y_{in}). (4) The purpose of the activation function is to determine if a particular neuron ought to be activated or not. This choice hinges on the significance of the neuron's input for the prediction process. (5) Bias alters the value produced by the summing junction.

Source: H2O.ai Team, 2025.

2.3.3 The Optimal ANN Model Design

The model network proposed in this paper was built as follows:

(1) Input layer: Includes five factors being Gender (2 inputs), Marital Status (5 inputs), Education (4 inputs), Residence (8 inputs), Work Status (5 inputs), and one covariate being Age (1 input); making a total of 25 input units.

(2) Proposed number of hidden layers = 2. (The two-hidden-layer model performed much better than a single hidden-layer model.)

(3) Number of units (Artificial neurons) in Hidden Layer 1 (Excluding the bias unit) = 9.

(4) Number of units (Artificial neurons) in Hidden Layer 2 (Excluding the bias unit) = 7.

a. Activation function for all neurons in hidden layers: The hyperbolic tangent (tanh) is a category of activation function that converts its input into a value ranging from -1 to 1. It is mathematically expressed as:

b. $$\tanh(Y_{in}) = \frac{e^{Y_{in}} - e^{-Y_{in}}}{e^{Y_{in}} + e^{-Y_{in}}} \quad (1)$$

c. Where "e" is Euler's number (approximately 2.718), and Y_{in} is the Summing Junction (Sum of the weighted inputs and bias of a neuron).

(5) Activation function for the two dependent variables: 'Heard of CRC' and 'Did not hear of CRC' is Identity. The identity activation function, also known as the linear activation function, is a type of activation function used in artificial neural networks (ANNs). It is defined as:

a. $$f(Y_{in}) = Y_{in} \quad (2)$$

b. This means that the output of the neuron is the same as its input. In other words, the function does not transform the input signal 'Y_{in}'; it simply passes it through unchanged.

2.3.4 The Training Process

The training algorithm follows standard backpropagation, which is a widely used algorithm for training ANNs (Hejase & Hejase, 1997; Surakhi & Salameh Walid, 2014; GeeksforGeeks, 2025c). It is a technique for effectively computing the gradient of the loss function concerning the network's weights and biases, allowing for the application of gradient-centric optimization methods such as gradient descent to adjust these parameters and reduce the error (GeeksforGeeks, 2025c).

The algorithm adheres to the following steps:

(1) Forward Pass: Generate predictions with existing weights, with the input data entering through the input layer. These inputs, along with their associated weights, are sent to hidden layers. In our suggested network featuring two hidden layers (H1 and H2), the output of H1 acts as the input to H2. A bias is added to the weighted inputs before applying the tanh activation function. At the output neuron, the inputs from the second hidden layer are summed (along with the bias), and the identity activation function delivers the classification probabilities.

(2) Loss Calculation: Determine the discrepancy between the anticipated and actual results.

(3) The Sum of Squares Error function (E) is calculated for a whole epoch (An epoch refers to one complete pass through the entire training dataset). During a single epoch, the neural network processes all training examples, and the sum of squares is:

$$\boldsymbol{E} = \sum(\mathbf{target\_output} - \mathbf{predicted\_output})^{\mathbf{2}} \qquad (3)$$

(4) Backward Pass: Compute gradients via backpropagation. The Backward Pass, or backpropagation, is a crucial step in training the neural network, following the forward pass and loss calculation. Its purpose is to compute the gradients of the loss function for each parameter (weights and biases) in the network. These gradients indicate how much each parameter should be adjusted to minimize the loss (Hakim, 2025).

(5) Parameter Update: Apply the Adam optimization step. The Adam optimization step involves calculating the first and second moments of the gradients, applying bias correction, and then updating the model parameters using an adaptive learning rate (Khare, 2023).

(6) Loop back to the Forward Pass to continue training the neural network. This iterative process repeats for a set number of epochs (one epoch is a pass through the training data) or until a convergence criterion is met, allowing the network to refine its weights and improve its performance.

(7) Validation: Evaluate performance on held-out validation set.

3. Results and Analysis

3.1 Data Exploration Results

The CRC awareness data distribution shows relatively balanced classes, with a slight preference towards "Yes" responses (656 aware of CRC and 563 unaware). Thus, the Dataset shows a good balance between awareness classes, and there is no significant class imbalance that requires specialized handling.

3.2 Training Performance

The data set was segmented into three separate sections: training, test, and validation sets. This separation maximizes the effectiveness of our neural network. The training dataset serves as the educational stage for the model. It's where the neural network attempts to recognize patterns in the data. The test set aids in assessing how effectively our model is acquiring knowledge. The validation set is utilized solely once the model has been completely trained. It is beneficial to evaluate how effectively our model functions on entirely new data (Codewave, 2024).

The ANN model demonstrates excellent training characteristics (see Table 1).

Convergence Analysis as shown in Table 2, where

(1) Training Accuracy: Converges to approximately 63.4% (51.3% for unaware cases and 73.5% for aware cases).

(2) Testing Accuracy: Converges to approximately 61.9% (46.2% for unaware cases and 77.4% for aware cases).

(3) Validation Accuracy: Achieves approximately 64.5% final accuracy with 60.7% for unaware cases and 68.3% for aware cases.

(4) Generalization: Validation accuracy exceeds training accuracy, indicating good generalization.

Table 1. Model Summary

Training	Sum of Squares Error	214.353
	% Incorrect Predictions	36.6%
Testing	Sum of Squares Error	24.014
	% Incorrect Predictions	38.1%
Holdout	% Incorrect Predictions	35.5%

Table 1 presents a "summary of the ANN outcomes by partition and in total. It includes the error, the relative error, or the proportion of incorrect predictions, and the duration of training" (IBM, 2019, p. 14).

Table 2. Classification: Comparing Observed Data to Predicted Data

Sample	**Observed: 'Heard of CRC'**	**Predicted: 'Heard of CRC'**		
		No	**Yes**	**Percent Correct**
Training	No	231	219	51.3%
	Yes	144	399	73.5%
	Overall Percent	37.8%	62.2%	63.4%
Testing	No	24	28	46.2%
	Yes	12	41	77.4%
	Overall Percent	34.3%	65.7%	61.9%
Holdout	No	37	24	60.7%
	Yes	19	41	68.3%
	Overall Percent	46.3%	53.7%	64.5%

3.3 Confidence Scoring Implementation

The prediction confidence is calculated using the distance from the decision boundary (Vemuri, 2020):

$$\textbf{Confidence score} = |p - 0.5| \times 2 \qquad (4)$$

Where p is the predicted probability given by the model neural network. The confidence level is categorized as:

- ✓ High: confidence score ≥ 0.6
- ✓ Medium: 0.2 ≤ confidence score < 0.6
- ✓ Low: confidence score < 0.2

3.4 Independent Variable Importance

In a neural network, independent variables (also known as input features or predictors) are assessed for their importance in predicting the output, or dependent variable. This importance is often quantified using various methods that analyze how much the network relies on each input to make accurate predictions (IBM, 2019).

Independent variable importance is determined by analyzing how much the model's output changes when the input (independent variable) is perturbed. This is typically done after the model has been trained and is often based on how much the error changes during training or testing. Figure 3 presents the normalized importance of the input features, where it is observed that the two most important features in CRC awareness are the place of residence, "Where the respondent lives," and the age. Educational level comes in third place, so this should be a motivation factor for both health and academic organizations to spread more CRC awareness campaigns among the educated communities (Hamza, Argaw, & Gela, 2021; A. Hejase et al., 2025).

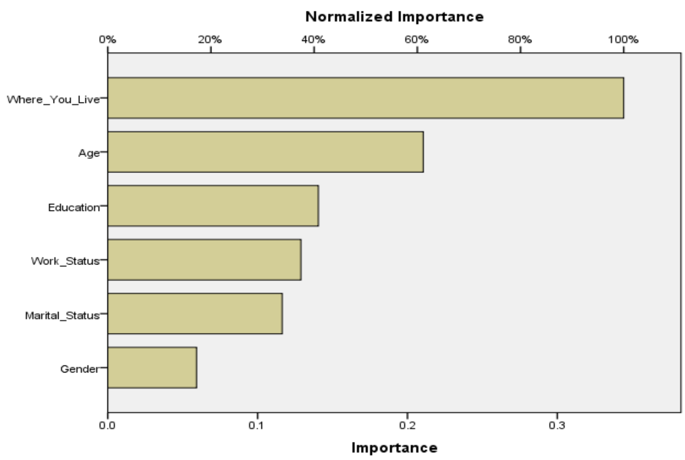

Figure 3. Individual Predictors' Importance in Determining the ANN Model

3.5 The Application Programming Interface for Real-time Predictions

The resulting ANN model implemented used an excel spread sheet to implement a user friendly application where the user may input the six predictors being: Age (one standardized input with a mean age of 26.98 years and a standard deviation of 9.69 years), gender (2 inputs, using one-hot encoding for male and female), marital status (5 inputs, using one-hot encoding for single, married, separated, divorced, and widowed), education (4 inputs, using one-hot encoding for higher studies, university, high school, and less than high school), region where they live (8 inputs, using one hot encoding for Akkar, Mount Lebanon, Bekka, North Lebanon, Dahie, Beirut, South Lebanon, and Shouf), and work status (5 inputs, using one-hot encoding for works, no-work, housewife, retired, and studying). The spreadsheet will perform all the calculations necessary to produce the two output probabilities: Being aware of CRC and not being aware.

The sequence of spreadsheet calculations goes as follows:

1. Use the 25 input predictors plus the bias to calculate the summing junction for each of the nine artificial neurons that belong to the first hidden layer (The weights for each input and bias are the optimal weights obtained in the proposed model).
2. Use the hyperbolic tangent activation function to find the output of each of the nine hidden neurons that belong to the first hidden layer.
3. The nine outputs of the first hidden layer, plus a bias input, are injected into the inputs of the seven artificial neurons that belong to the second hidden layer (The weights for each input and bias are the optimal weights obtained in the proposed model).
4. Use the hyperbolic tangent activation function to find the output of each of the seven hidden neurons that belong to the second hidden layer.
5. The seven outputs of the second hidden layer, plus a bias input, are injected into the inputs of the two artificial neurons that belong to the output layer (The weights for each input and bias are the optimal weights obtained in the proposed model).
6. Use the identity activation function to find the output of each of the two output neurons that belong to the output layer. The two output neurons provide the probabilities for being aware of CRC and for not being aware.
7. Finally, the higher probability is chosen as the prediction outcome, and the output message will be '*Person Aware*' or '*Person Unaware*.'

Table 3 illustrates a screenshot of the Excel API interface, and Tables 4 to 6 are sample screenshots presenting the optimal weights that correspond to the nine neurons of the first and second hidden layers, and the outputs. Finally, Figure 4 depicts the ANN model for CRC Awareness.

Table 3. The Excel Interface API for CRC Awareness Prediction Based on Demographic Data

To predict CRC awareness fill the following demographic data:	
Age	50
Gender	Female
Marital Status	Single
Education	Less Than High School
Region Where you live	Beirut
Work Status	Housewife
Prediction	Person Aware

Table 4. The model's optimal weights for the nine neurons of the first hidden layer.

	Predictor	H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)
Input Layer	(Bias)	0.14	0.326	0.078	-0.261	0.067	-0.488	0.274	-0.017	-0.341
	[Gender=0]	0.625	-0.068	0.299	0.159	0.269	-0.441	-0.231	0.126	0.17
	[Gender=1]	-0.209	0.385	0.069	-0.281	-0.202	-0.303	-0.295	-0.193	0.42
	[Marital_Status=1]	-0.019	0.422	0.101	0.331	0.567	-0.274	0.565	0.175	0.192
	[Marital_Status=2]	0.055	0.193	-0.038	-0.243	-0.424	-0.272	0.411	-0.844	-0.258
	[Marital_Status=3]	-0.202	0.18	0.197	-0.095	0.43	0.049	-0.443	-0.148	-0.195
	[Marital_Status=4]	-0.43	0.097	0.083	-0.238	0.021	-0.336	0.018	-0.331	0.103
	[Marital_Status=5]	-0.263	0.08	-0.434	0.285	0.066	-0.351	0.006	-0.11	-0.411
	[Education=1]	0.101	-0.278	-0.22	0.406	-0.27	0.077	0.276	0.053	-0.325
	[Education=2]	0.189	0.155	-0.059	-0.071	0.455	-0.407	-0.338	0.385	0.399
	[Education=3]	0.248	0.377	0.413	-0.061	0.312	-0.058	-0.125	-0.385	0.321
	[Education=4]	-0.279	0.229	0.142	0.236	-0.089	-0.087	-0.416	-0.534	0.194
	[Where_You_Live=1]	0.102	-0.207	-0.42	0.191	-0.178	-0.35	0.25	-0.419	-0.338
	[Where_You_Live=2]	-0.339	0.487	0.324	-0.157	0.436	0.276	0.054	0.156	0.298
	[Where_You_Live=3]	0.288	0.443	0.503	-0.48	0.452	-0.427	0.155	0.286	-0.152
	[Where_You_Live=4]	0.384	-0.727	-0.098	0.091	-0.298	0.079	-0.127	-0.093	0.764
	[Where_You_Live=5]	-0.048	0.334	-0.255	0.293	-0.177	-0.256	-0.554	0.378	0.545
	[Where_You_Live=6]	0.348	0.285	-0.066	0.1	0.288	0.181	0.242	-0.161	-0.436
	[Where_You_Live=7]	0.34	0.038	-0.639	-0.148	0.302	-0.28	0.535	-0.337	0.1
	[Where_You_Live=8]	0.316	0.398	0.461	0.374	-0.477	0.446	0.462	0.573	-0.354
	[Work_Status=1]	0.3	0.04	-0.43	-0.365	-0.262	-0.207	-0.145	-0.288	0.151
	[Work_Status=2]	-0.203	0.167	0.199	-0.392	0.46	-0.25	0.305	-0.106	-0.113
	[Work_Status=3]	-0.146	0.458	0.299	-0.002	-0.095	0.01	0.029	-0.32	0.161
	[Work_Status=4]	-0.179	-0.32	-0.518	-0.125	-0.071	0.254	-0.38	0.358	0.407
	[Work_Status=5]	0.475	0.327	-0.348	0.27	0.122	0.456	-0.513	0.352	0.356
	Age	-0.859	-0.245	0.078	-0.491	0.087	0.384	-0.155	0.675	-0.138

Table 5. The Model's Optimal Weights for the Nine Neurons of the Second Hidden Layer.

	PREDICTOR	H(2:1)	H(2:2)	H(2:3)	H(2:4)	H(2:5)	H(2:6)	H(2:7)
HIDDEN LAYER 1	(Bias)	0.507	0.791	-0.023	0.127	-0.184	-0.109	0.194
	H(1:1)	-0.162	-0.262	0.026	-0.048	-0.164	0.375	0.589
	H(1:2)	-0.325	-0.22	0.085	0.788	-0.297	-0.071	0.293
	H(1:3)	0.155	0.422	-0.132	0.199	-0.059	0.38	0.241
	H(1:4)	-0.234	-0.17	-0.397	-0.139	0.066	-0.092	-0.322
	H(1:5)	0.257	0.328	-0.141	0.238	-0.078	0.201	-0.437
	H(1:6)	-0.352	-0.041	-0.053	0.027	-0.086	0.258	0.025
	H(1:7)	-0.024	-0.199	-0.294	-0.164	-0.123	0.559	-0.163
	H(1:8)	0.507	0.383	0.401	0.026	-0.005	0.585	0.672
	H(1:9)	0.366	0.431	0.008	-0.099	-0.269	-0.554	0.303

Table 6. The Model's Optimal Weights for the Outputs

	Predictor	[Heard_of_CRC=0]	[Heard_of_CRC=1]
Hidden Layer 2	(Bias)	0.587	0.383
	H(2:1)	0.413	-0.076
	H(2:2)	0.279	-0.513
	H(2:3)	-0.487	0.234
	H(2:4)	-0.309	0.524
	H(2:5)	0.374	0.06
	H(2:6)	-0.362	0.291
	H(2:7)	-0.351	0.529

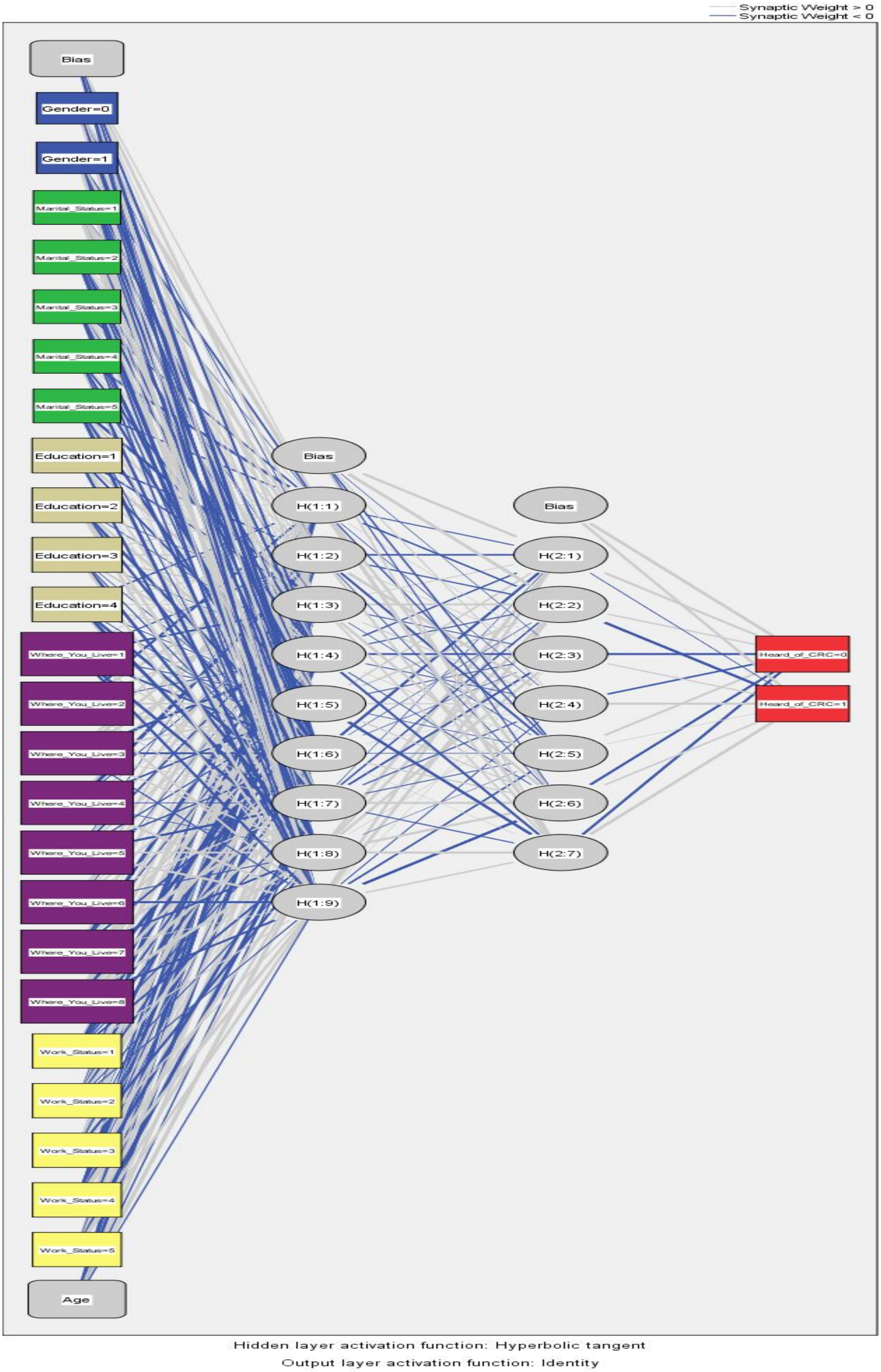

Figure 4. The Resultant ANN Model for CRC Awareness

4. Conclusions

Neural network implementation (Model is presented in Figure 4) provides a comprehensive, production-ready solution for CRC awareness prediction using modern deep learning techniques. The neural network architecture demonstrates the appropriate complexity for the demographic prediction domain, while incorporating state-of-the-art regularization techniques to ensure robust generalization.

4.1 Key Achievements

Technical Accomplishments:

1. Successful binary classification with 64.5% validation accuracy.
2. Robust data preprocessing pipeline with feature scaling and validation.
3. Extensible architecture that supports future enhancements.
4. Comprehensive model analysis and visualization capabilities.

4.2 Impact and Applications

The implemented solution serves multiple purposes:

1. Healthcare Research: Provides insights into demographic factors affecting CRC awareness.
2. Public Health Planning: Enables targeted awareness campaigns based on demographic profiles.
3. Educational Tool: Demonstrates best practices in machine learning model deployment.
4. Foundation Platform: Serves as an extensible framework for similar healthcare prediction tasks.

4.3 Strengths and Limitations

Our study has strengths and limitations. First, it is the first study to build a prediction model for CRC awareness. This implementation establishes a solid foundation for healthcare awareness prediction tasks and demonstrates the industry's best practices in machine learning model development and deployment. Moreover, the ANN model provides valuable insights to help guide the implementation and improvement of such public health strategies. However, there are some limitations, such as the modest validation accuracy of approximately 64.5% final accuracy, with 60.7% for unaware cases and 68.3% for aware cases. These figures could be improved with higher sample sizes and better control of social desirability bias. Additionally, future work needs to go beyond demographics by improving data collection to include psychosocial and behavioral predictors.

References

Abreu Lopez, B.A., Pinto-Colmenarez, R., Caliwag, F.M.C., Ponce-Lujan, L., Fermin, M.D., Granillo-Cortés, A.V., Mejía-Martínez, A.G., Zepeda-Martinez, I.G., & León, F.G. (2024, October 6). Colorectal Cancer Screening and Management in Low- and Middle-Income Countries and High-Income Countries: A Narrative Review. *Cureus Journal of Medical Science*, *16*(10), e70933. https://doi.org/10.7759/cureus.70933

Altıntaş, O., Aksoy, M., Ünal, E., Akgöl, O., & Karaaslan, M. (2019). Artificial Neural Network Approach for Locomotive Maintenance by Monitoring Dielectric Properties of Engine Lubricant. *Measurement*, *145*, 678–686. https://doi.org/10.1016/j.measurement.2019.05.087.

American Cancer Society. (2024, January 29). *Guideline for Colorectal Cancer Screening*. Retrieved July 15, 2025, from https://www.cancer.org/cancer/types/colon-rectal-cancer/detection-diagnosis-staging/acs-recommendations.html

CDC. (2025, March 11). Colorectal Cancer Awareness. *Centers for Disease Control and Prevention*. Retrieved July 15, 2025, from https://www.cdc.gov/cancer/features/colorectal-cancer.html

Chehimi, G.M., & Hejase, H.J. (2024). Exploring the Impact of Bilingualism in Early Life on Foreign Language Learning for University Students in Lebanon, Theoretical Foundations: Part I. *Journal for the Study of English Linguistics*, *12*(1), 1-31. https://doi.org/10.5296/jsel.v12i1.21684

Cleveland Clinic. (2022, November 14). *Colorectal (Colon) Cancer*. Retrieved July 25, 2025, from https://my.clevelandclinic.org/health/diseases/14501-colorectal-colon-cancer

Codewave. (2024, October 15). Steps to Create and Develop Your Own Neural Network. *Codewave Technology*. Retrieved July 31, 2025, from https://codewave.com/insights/how-to-develop-a-neural-network-steps/

Eka Rusdi Antara, G. (2024). Challenges of Colorectal Screening in Developing Countries. In Jindong Chen (Ed.) *Advances in Diagnosis and Therapy of Colorectal Carcinoma* (Chapter 4). IntechOpen. https://doi.org/10.5772/intechopen.1004062

GeeksforGeeks. (2025a, July 12). *Artificial Neural Networks and Their Applications*. Retrieved July 31, 2025, from https://www.geeksforgeeks.org/artificial-intelligence/artificial-neural-networks-and-its-applications/

GeeksforGeeks. (2025b, July 23). *Weights and Bias in Neural Networks*. Retrieved July 31, 2025, from https://www.geeksforgeeks.org/deep-learning/the-role-of-weights-and-bias-in-neural-networks/

GeeksforGeeks. (2025c, July 01). *Backpropagation in Neural Networks*. Retrieved July 31, 2025, from https://www.geeksforgeeks.org/machine-learning/backpropagation-in-neural-network/

H2O.ai Team. (2025). What Is Neural Network Architecture? *H2O Wiki*. Retrieved July 25, 2025, from https://h2o.ai/wiki/neural-network-architectures/

Hadoua, A., Belaadi, A., Ghernaoutb, D., & Mukalazid, H. (2025). Forecasting the Thermal Degradation Depending on the Kinetics of Dracaena Draco Lignocellulosic Fibers Using an Artificial Neural Network. *Journal of Natural Fibers*, *22*(1), 2531368. https://doi.org/10.1080/15440478.2025.2531368

Hakim, M. (2025, March 1). Backpropagation in Neural Networks. *Medium*. Retrieved July 25, 2025, from https://medium.com/@mudassar.hakim/backpropagation-in-neural-networks-b7f7dd595de7

Hamza, A., Argaw, Z., & Gela, D. (2021). Awareness of Colorectal Cancer and Associated Factors among Adult Patients in Jimma, South-West Ethiopia: An Institution-Based Cross-Sectional Study. *Cancer Control*, *28*, 10732748211033550. https://doi.org/10.1177/10732748211033550

Hassibi, K. (2016, October 28). Machine learning vs. traditional statistics: Different philosophies, different approaches. *Data Science Central* (I. TechTarget, Editor). Retrieved July 25, 2025, from https://www.datasciencecentral.com/machine-learning-vs-traditional-statistics-differentphilosophi-1/

Hejase, A.J., & Hejase, H.J. (2013). *Research Methods, A Practical Approach for Business Students* (Second ed.). Philadelphia, PA, USA: Masadir.

Hejase, A.J., Hejase, H.J., Kassem, R.H., & Hejasebazzi, A.A. (2025, August). Colorectal Cancer (CRC) Awareness in Lebanon among the Young, Highly Educated Population. *Research in Health Science*, *10*(3), 81-110. https://doi.org/10.22158/rhs.v10n3p81

Hejase, H.J., & Hejase, A.J. (1997). Using Neural Networks to Model experimental data obtained from Torque Magnetometer experiments on Cr-thin films. *Proceedings of the 2nd LAAS International Conference on Computer Simulation*, Beirut, Lebanon. Retrieved July 31, 2025, from https://www.researchgate.net/publication/236863199_Using_Neural_Networks_to_Model_experimental_data_obtained_from_Torque_Magnetometer_experiments_on_Cr-thin_films

Hejase, H.J., Fayyad-Kazan, H., Hejase, A.J., Moukadem, I., & Danach, K. (2023a). Needed MIS Competencies to the Job Market: Students' Perspective. *British Journal of Multidisciplinary and Advanced Studies: Business and Management Sciences*, *4*(5), 120-162. https://doi.org/10.37745/bjmas.2022.0324

Hejase, H.J., Rkein, H., Hamdar, B., and Hejase, A.J. (2023b). Needed Accounting Competencies to the Job Market. *British Journal of Multidisciplinary and Advanced Studies: Business and Management Sciences*, *4*(5), 1-17. https://doi.org/10.37745/bjmas.2022.0289

IBM. (2019). *IBM SPSS Neural Networks 26*. Retrieved July 15, 2025, from https://www.ibm.com/docs/en/SSLVMB_26.0.0/pdf/en/IBM_SPSS_Neural_Network.pdf

Khare, Y. (2023, December 24). What is Adam Optimizer and How to Tune its Parameters in PyTorch? *Analytics Vidhya*. Retrieved July 15, 2025, from https://www.analyticsvidhya.com/blog/2023/12/adam-optimizer/

Li, Q., Geng, S., Luo, H., Wang, W., Mo, Y.-Q., Luo, Q., Wang, L., Song, G.-B., Sheng, J.-P., & Xu, B. (2024, October 7). Signaling pathways involved in colorectal cancer: Pathogenesis and targeted therapy. *Sig Transduct Target Ther*, *9*, 266. https://doi.org/10.1038/s41392-024-01953-7

Masoudi, O.A., & Hejase, H.J. (2023). Needed Current Characteristics of a Good Iranian Graduate in Journalism and Media Studies. *Asian Business Research*, *8*(2), 1-13. https://doi.org/10.20849/abr.v8i2.1368

Nemer, H., Hejase, A., Hejase, H., Othman, M., Chawraba, M., & Trad, M.-A. (2016). Colorectal Cancer: Exploring Awareness in Lebanon. *The Journal of Middle East and North Africa Sciences*, *2*, 10-21. https://doi.org/10.12816/0032694

Pankratz, V.S., Kanda, D., Kosich, M., Edwardson, N., English, K., Adsul, P., & Mishra, S.I. (2024). Racial and Ethnic Disparities in Colorectal Cancer Incidence Trends Across Regions of the United States From 2001 to 2020 - A United States Cancer Statistics Analysis. *Cancer Control*, *31*, 10732748241300653. https://doi.org/10.1177/10732748241300653

Peace, P., Chris, J., & Victor, L. (2024, November). Data Preprocessing for AI Models. *ResearchGate*. Retrieved July 15, 2025, from https://www.researchgate.net/publication/385707249_Data_Preprocessing_for_AI_Models

Pisa, I., Santín, I., Lopez Vicario, J., Morell, A., & Vilanova, R. (2019). Data Preprocessing for ANN-based Industrial Time-Series Forecasting with Imbalanced Data. *Proceedings of the 2019 27th European Signal Processing Conference* (EUSIPCO), (pp. 1-5). https://doi.org/10.23919/EUSIPCO.2019.8902682

Rammal, H., Hejase, H.J., & Hazimeh, H. (2024). Metaverse Technology and Its Impact on the Evolving Landscape of Communication and Media: A Future Outlook for Lebanese Satellite Channels. *Saudi J. Humanities Soc Sci*, *9*(3), 92-117. https://do.org/0.36348/sjhss.2024.v09i03.001

Rammal, H.E., Hejase, H.J., & Hejase, A.J. (2025, July). Assessment of Academic Art Stakeholders' Stance Toward Artificial Intelligence: A Study From Lebanon. *Asian Journal of Social Science Studies*, *10*(3), 1-23. https://doi.org/10.20849/ajsss.v10i3.1517

Ranganath, R., Lujaina, S. S.-m., Al Saidi, A., Al Dujaili, A. H., & Mahmood, M. A.-F. (2022). Knowledge and Awareness of Colorectal Cancer among University Students in Oman. *J Health Sci Med Res*, *40*(2), 215-222. http://dx.doi.org/10.31584/jhsmr.2021835

Rawla, P., Sunkara, T., & Barsouk, A. (2019). Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Przeglad Gastroenterologiczny*, *14*(2), 89-103. https://doi.org/10.5114/pg.2018.81072

Stewart, M. (2019, March 24). The actual difference between statistics and machine learning. *Medium* [Towards Data Science (TDS) Archive, (March)]. Retrieved July 25, 2025, from https://medium.com/data-science/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3

Surakhi, O.M., & Salameh Walid, A. (2014). Enhancing the Performance of the Backpropagation for Deep Neural Networks. *International Journal of Computers & Technology*, *13*(12), 5274-5285. https://doi.org/10.24297/ijct.v13i12.5279

Turco, C., Funari, M.F., Teixeira, E., & Mateus, R. (2021). Artificial Neural Networks to Predict the Mechanical Properties of Natural fibre-Reinforced Compressed Earth Blocks (CEBs). *Fibers*, *9*(12), Art. No 12. https://doi.org/10.3390/fib9120078

Van Altena, W. (2021, March). To know is to act': new campaign seeks to raise colon cancer awareness. *Inside Oncology*. Retrieved July 31, 2025, from https://www.bjmo.be/to-know-is-to-act-new-campaign-seeks-to-raise-colon-cancer-awareness/

Vemuri, N. (2020). *Scoring Confidence in Neural Networks*. (Master of Science Research Project), Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, California, USA.