

# Predicting Human Gender through Population Data Using Artificial Neural Networks

Amadi, Mirabel Wucheckukwu<sup>1</sup>, Ojekudo, Nathaniel Akpofure<sup>2</sup>

P.G. Student, Department of Computer Science, Ignatius Ajuru University of Education, Port Harcourt,  
Rivers State, Nigeria<sup>1</sup>

Senior Lecturer, Department of Computer Science, Ignatius Ajuru University of Education, Port Harcourt,  
Rivers State, Nigeria<sup>2</sup>

**ABSTRACT:** Gender identification is a basic requirement when it comes to Profiling an individual. It is one clear way information can be used effectively, but not all individuals will provide this information publicly, this makes it very difficult to profile individual data with incomplete details, and when an individual is hit by a disastrous or any kind of traumatic situation the propensity to which people are eager to know the gender of the individual concerned becomes a vital ingredient of the subject. Disaggregating statistical data on individual characteristics by sex helps to provide evidence of gender differences as it also strengthens and improves the entire statistical system. However, this research focuses on a fast and efficient method of predicting human gender by computing/analyzing population data into different gender groups using artificial neural network. The final decision is made by employing validation from a binary prediction decision “male/female” (with male indicating a 1 and 0 indicating a female) method based on experimental results obtained from comparing the computed mean with the total population figure using a mean threshold operator  $<$ ,  $>$  or its equivalent ( $=$ ). The proposed method is better in terms of speed and accuracy.

**KEYWORDS:** Gender Prediction, Forensic Science, Population Data, Artificial Neural Network.

## I. INTRODUCTION

One sure way to use information is to study the relation of conditions by identifying potential basic inclusionary criteria that have to be met, such as gender. But not all persons will provide this information publicly, thus it makes user profiling of data very difficult. Scarcity of relevant, reliable and correctly analyzed gender data is one major challenge that has kept policy-makers and development planners bound, as this difficulty is not new to Non-Governmental Organisations (NGOs) who carry out needs assessments, also university researchers who do surveys, and as well as governments that conduct census. When these data are not analyzed properly or presented in a format that is easy to use or accessible to the public or concerned individual making use of it, it becomes a major challenge. Disaggregating statistical data on individual characteristics by sex will help to provide evidence of gender differences as it also strengthens and improves the entire statistical system. This is because Women and Men have different roles in society, different needs and interests, different access to and control over resources. Thus, national statistics that do not reflect these differences are insufficient and potentially misleading. UNFPA, [1].

Furthermore, when individuals are hit by a catastrophic or any kind of traumatic situation, the sex of the person responsible (both the victim and the culprit) becomes a vital ingredient of the subject. As this creates an alarming rate at the propensity to which people are eager to know the gender of the concerned individual. As such, forensic or Medio-legal investigation becomes the first line of action to determine the victim Mumtaz and Kun [2]. The sex, age, stature and race of an individual (either living, dead or skeleton) is examined to determine his identity

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 10, October 2019

during forensic investigations. Larson et. al., [3]. It is also noted that the skull and pelvis areas of humans are the most common and popular parts for sex determination. Pinto et. al., [4].

## 1.1. Artificial Neural Networks (ANN)

Artificial neural networks represent an analogy to human thinking, a simplified imitation of the principle of human brain. They are described as the black box with unknown inner structure and known outputs Tereza, [5]. Artificial neural networks usually work in two stages. In the first stage the network acts as an inexperienced man, learning to set its parameters to correspond to the required network topology. In the second stage the network already independently transforms inputs to outputs on the basis of knowledge obtained in the first stage. Each neural network consists of defined layers, input layers, hidden layers and output layers of which each has several highly interconnected computational units called nodes or neurons and the learning method and the knowledge acquisition process. When applying ANN to prediction problems, the number of independent variables determines the number of input nodes and the predicted variables specifies the number of output nodes. Fendigilet. al., [6]. Furthermore, can identify patterns between dependent and independent variables in datasets. Its principal strengths are pattern recognition as well as optimization of large-scale problem. DeTienne and DeTienne [7]. Has advantage of effective interaction with data discontinuities, outliers missing data and nonlinear transformations. Restriction of number of hidden neurons that hinders its function and complex computation are some of its disadvantages.

## II. RELATED WORK

Gender predictions have been effectively predicted in many research areas such as forensic investigations, health image signal processing and social media user just to name a few.

About 600 persons were investigated to determine their gender, considering their hand dimension, length and breadth. The ratio between their index fingers were also considered. Model result showed that women have higher ratio of index and ring fingers than men. Ibrahim et. al., [8].

A dataset consisting of 250 men and 250 women of upper Egyptian adults over 18 years of age, were investigated to determine their gender, by calculating their hand dimensions, index and ring finger ratio. Their model showed that the average male hand is 1.3cm larger than the female. Aboulet. al., [9].

Military Sinus Volumes (MVS) were used to determine the gender of individuals by using 3D reconstruction. A dataset consisting of three hundred and three (303) CT scan images were considered and result suggested that MSV establishment can be useful in determining individual's sex. Thomas et. al., [10].

200 patients' CT scan skull was investigated, estimating their scapular diameters. Estimation was carried out by analyzing the scapular of each patient after 3D post processing reconstructions. Their model result indicated a high level of accuracy as 88% of the calibration sample proved correct as it identified 9 out of 10 testing sample. Giurazza, et. al., [11].

The measurement values of foramen magnum in relation to the determination of biological gender was investigated on 250 adult samples using CT scans. Their model result indicated a high level of accuracy as adult samples were matured enough to point out significant clues which helped to differentiate gender related features in male and female adults. Edwards et. al., [12].

Biometric tracts found in the hand was used to determine individual gender using Convolutional Neural Networks. Palm and normal parts of hands were the images used on both sides. Their result showed good accuracy not only on palm side but also on dorsal side of the hand. Afifi, et. al., [13].

A GDCNN (Gender Determination with Convolutional Neural Networks) technique, where the left hand radio-graphs of the children between a wide range of ages between 1 month to 18 years was examined to determine their gender, this technique was however, the first of its kind. A Class Activation Mapping (CAM) was used further to identify the area of the attention. Model result suggested an accuracy level as high as 98%, which was very convincing by taking into account the incompletely grown skeleton of the children. The attention observed with CAM discovers that the lower part of the hand around carpals (wrist) is more important for child gender determination. Mumtaz and Kun [2].

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 10, October 2019

Age group, gender and personality traits were predicted on Facebook posts using n-gram and LDA topics as features to obtain an accuracy level of 91.9% for gender and a regression coefficient of 0.84 for age. A word cloud for people of divers age and gender was also generated which showed a distinct difference between people of different ages and features. Schwartz et. al., [14].

A computational and dynamic face aging model was developed to classify human gender. Research was based on a geometric ratio and skin wrinkle analysis. Study result showed 100% classification accuracy when tested on a database of only 47 high resolution face images of babies, young and middle-aged adults. Suo, et. al., [15].

A mechanism for machine classification of human gender with automatic confidence and its application to gender classification was proposed. This classified human gender based on facial extraction features using Support Vector Machine (SVM). Study result was very efficient and accurate. Zheng and Lu [16].

A human gender and age classification system to classify facial images into different gender and age group using Feed Forward Artificial Neural Networks (FFANN) was proposed. A three sigma control limits in the finer level was the final decision made after validation. Model was better in terms of speed and accuracy as a single frontal human face with two gender groups and with three different age groups were detected successfully with 95% success rate.

## III. METHODOLOGY

### 3.1. Gender Determination with ANN

Dataset used prediction is the Nigeria population size by gender, consisting of 934.37 Male and 911.86 female (in Millions) ranging from a period of 2007 to 2017, was sourced from the ([www.sttista.com](http://www.sttista.com)) website, was used for prediction experiment,

**Table 3.1: Sample of Original Dataset, (Source: [www.sttista.com](http://www.sttista.com))**

S/N	Year	Male (in Millions)	Female (in Millions)
1	2017	94.16	96.73
2	2016	91.76	94.23
3	2015	89.41	91.77
4	2014	87.11	89.35
5	2013	84.84	86.98
6	2012	82.63	84.67
7	2011	80.47	82.40
8	2010	78.37	80.20
9	2009	76.33	78.07
10	2008	74.35	75.99
11	2007	72.43	73.98

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 10, October 2019

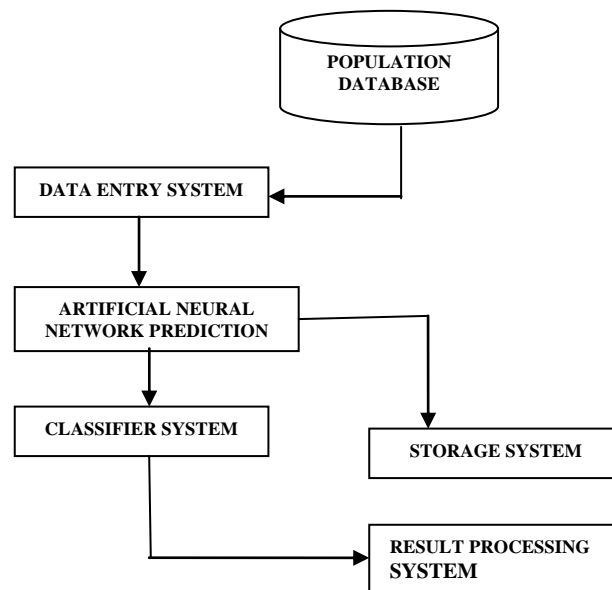
Table 3.1 shows the yearly statistics of dataset used for experimentation. The male and female population figures displayed in Millions is used as input for predicting human gender, with the goal is to exploit the great potentials of Artificial Neural Network (ANN) technology in the context of gender prediction.

**Table 3.2: Sample Pre-processed Dataset (source: [www.ststista.com](http://www.ststista.com))**

S/N	Year	Male (in Millions)	Female (in Millions)	Total (in Millions)
1	2017	96.73	94.16	190.89
2	2016	94.23	91.76	185.99
3	2015	91.77	89.41	181.18
4	2014	89.35	87.11	176.46
5	2013	86.98	84.84	171.82
6	2012	84.67	82.63	167.30
7	2011	82.40	80.47	162.87
8	2010	80.20	78.37	158.57
9	2009	78.07	76.33	154.40
10	2008	75.99	74.35	150.34
11	2007	73.98	72.43	146.41
<b>Total</b>		<b>934.37</b>	<b>911.86</b>	<b>1, 846.23</b>

Table 3.2 shows the pre-processed dataset after an initial analysis and experimentation was conducted on the dataset in table 3.1. With the population figures for male and female as well as the total population size was considered as our (target) output to the system.

### 3.2. System Architecture



**Figure 3.1: System Architecture of the Artificial Neural Network Prediction System**

Figure 3.1 illustrates the System Architecture of the Artificial Neural Network Prediction System and thus involves of a Population Database that houses all the variables considered for prediction in the Artificial Neural Network system is created, this includes; the year, male and female population figures as shown in the dataset displayed in table 1. However, after an initial analysis and experimentation on the dataset, (data pre-processing). The population figures for male and female were however, used as inputs to the system, while the total population size was considered as our (target) output to the system.

The system allows the entrance of only two variables (male and female population figures) that are considered as inputs to the system. As it is a convenient user interface that ensures the interaction between a user and the Artificial Neural Network system.

Predictions are made in the system by computing the output (target) based on the parameters considered as inputs to the system after analysis and experimentation have been performed on the dataset. Thus, the output is computed using the equation:

$$\text{Input } G_i = \text{Target element input}$$

Input is transformed to get our target element, as thus;

$$G_i(\text{transform}) = \begin{cases} G_i > \mu G_i & = 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

where,

$\mu G_i$  = Mean Threshold Operator (MTO).

$$\text{Thus, } \mu G_i = \frac{\sum G_i}{n} \quad (3.2)$$

where,  $n$  is the total number of elements.

and  $\sum G_i$  is the Total Number of Target Element .

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)  
 Vol. 8, Issue 10, October 2019

Suppose the dataset given in table 2

From the dataset given in table 2, MTO is computed as thus;

The Mean target (output) =  $\frac{\text{Total Number of Target Element}}{\text{Total Number of Input Element}}$

$$= \frac{1846.23}{11}$$

$$= 167.84$$

Thus,  $MTO = 167.84$

Input is transformed to get our target element, as thus; the computed mean (1846.23) is used to conditionally generate patterns of 1's (Male) or 0's (Female) using a threshold operator (<, > or its equivalent (=)).

$G_i$ , is thus computed as; compare  $MTO$  with each target element in the dataset given in table 2, this generates an output class representation in binary form as thus, if  $MTO$  is greater than the target element we have a 1, which indicates Male Gender, but if  $MTO$  is less than the target element we have a 0, and that indicates Female Gender. Hence our output class representation in binary form is given as thus;

**Table 3.3: Output Class Representation**

S/N	$G_i$ (transform)
1	0
2	0
3	0
4	0
5	0
6	1
7	1
8	1
9	1
10	1
11	1

Table 3.3 shows our output class representation in binary from experimentation conducted using the binary prediction decision. As thus, if  $MTO$  is greater than the target element we have a 1(Male Gender) but if  $MTO$  is less than the target element we have a 0, and that indicates Female Gender.

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 10, October 2019

### 3.2.1. Computation of Individual Target Class Probability

We compute the individual target class probability using the equation

$$\text{Probability} = \frac{\text{Probability of Occurrence of an event}}{\text{Total Number of Event}} \quad (3.3)$$

Probability of Male class is  $6/11 = 0.55$

Probability of Female class is  $5/11 = 0.45$

Predictions generated by the Artificial Neural Network system are stored in a suitable database file for later access. The Artificial Neural Network prediction systems performance is evaluated on the basis of the percentage classification accuracy. The percentage accuracies are estimated considering the True-Positive (TP) and the True-Negative (TN) state for each prediction compared to the actual (target). In addition, if the prediction is not a True-Positive (TP), then it is a False-Positive, and if the prediction is not a True-Negative (TN), then it is a False-Negative (FN).

Where;

TP = Total number of true-positive

TN = Total number of true-negative

FP = Total number of false-positive

FN = Total number of false-negative

## IV. EXPERIMENTAL RESULTS

Experimental results tabulated in table 3.3, in section 3 shows how efficient the input variables were transformed to obtain our target element, as the computed mean was used to conditionally generate patterns of 1's (Male) or 0's (Female) using a threshold operator (<, > or its equivalent (=)) this was achieved by comparing *MTO* (computed mean) with each target element (total population figure) in the dataset given in table 2, as thus, if *MTO* is greater than the target element we have a 1, which indicates Male Gender, but if *MTO* is less than the target element we have a 0 and that indicates female gender. Table 3.1, also illustrates that from our experimental result, 1 to 5 predicted female genders while 6 to 11 predicted male genders.

## V. CONCLUSION

A fast and efficient human gender prediction method is proposed to analyze population data, into different gender groups using artificial neural network. The goal is to exploit the great potentials of Artificial Neural Network (ANN) technology in the context of gender prediction. The final decision is made by employing validation from a binary prediction decision "male/female" (with male indicating a 1 and 0 indicating a female), based on experimental results obtained from comparing the computed mean with the total population figure using a mean threshold operator <, > or its equivalent (=). The proposed method is better in terms of speed and accuracy. Fuzzy logic approach can be employed for further study as it will consider other classification features beyond gender in order to give a more precise judgement.

## REFERENCES

- [1] UNFPA, "Methodological Guidelines for Gender Analysis of National Population and Housing Census. *Technical Division of the United Nations Population Funds (UNFPA)*, 2018.
- [2] Mutaz, A., K. and Kun, H. "Child Gender Determination with Convolutional Neural Networks on Hand Radio-Graphs", *arXiv Preprint arXiv:1811.05180*, 2018.
- [3] Larson, B. D., Chen, M. C., Lungren, M. P., Halabi, S. S., et al., "Performance of Deep Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radio Graphs", *Radiology*, No. 287, 2017
- [4] Pinto, S., Urbanova, P., and Roberto, M. C. J., "Two-Dimensional Wavelet Analysis of Supraorbital Margins of the Human Skull for Characterizing Sexual Dimorphism", *IEEE Transactions On Information Forensics and Security*, vol. 11, No. 7, 2016



# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 10, October 2019

- [5] Tereza, S., "A Suitable Artificial Intelligence Model for Inventory Level Optimization", *Trends Economics and Management*. Pp. 45-55, vol. 25, No. 1, 2016.
- [6] Fenddigil, T., Onut, S., and Kahraman, C., "A Decision Support System for Demand Forecasting with Artificial Neural Networks and Neuro-Fuzzy Models", *A comparative Analysis of Expert Syst. Appl.* (6697– 6702) Vol 36, 2009
- [7] DiTienne, K. B., and DiTienne, D. H., "Neural Networks in Strategic Marketing: Exploring the Possibilities." *Journal of Strategic*. Pp. 1-12, Vol 29, 2015.
- [8] Ibrahim, M. A., Khalifa, A. M., Hagra, A., and N., I. Alwakid, N. I., "Sex Determination from Hand Dimensions and Index/ Ring Finger Length Ratio in North Saudi Population", *Egyptian Journal of Forensic Sciences*, pp. 435-444, Vol 42016.
- [9] Aboul-Hagag, K. E., Mohammed, S. A., Hilal, M. A., and Mohammed, E. A., "Determination of Sex from Hand Dimensions and Index/Ring Finger Ratio in Upper Egyptian," *Egyptians Journal of Forensic Sciences*, pp. 80-86, vol. 1, 2011
- [10] Thomas, R., Justin, M., Macini, M., and Dessi, P., "Sex Estimation from Human Cranium: Forensic and Anthropological Interest of Military Sinus Volumes." *Journal of Forensic Sciences*, pp. 805-808 vol. 6 Issue 3, 2019
- [11] Giurazza, F., Schena, E., Del, V. R., Cazzato, R. L., Mutato, L., et. al., "Sex Determination From Scapular Length Measurements by CT Scans Images in a Caucasian Population", *35<sup>th</sup> Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, (1632-1635)*, 2013
- [12] Edwards, k., Viner, M., Schweitzer, W., and Thali, M., "Sex Determination from Foramen Magnum", *Journal of Forensic Radiology and Imaging*, pp.186-192, Vol 1, No.4, 2010
- [13] Afifi, M., "Gender Recognition and Biometric Identification Using a Large Dataset of Hand Images", arXiv Preprint arXiv.1711.04322
- [14] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Diurzynski, M., Stillwell, D., and Stelgman, M. E., "Personality of Social Media. *The Open Vocabulary Approach*", *PLSONE*, Vol 8, Iss 9, 2013
- [15] Suo, J., Zhu, S. C., and Shan, S., "A Compositional and Dynamic Model for Face Aging", *IEEE Transactions and Pattern Analysis and Machine Intelligence*, Vol. 32, Iss 3, 2010
- [16] Zheng, J., and Lu, B. K., "A Support Vector Machine Classifier with Automatic Confidence and its Application to Gender Classification", *International Journal of Neuro-computing*, pp. 1926-35, Vol 74, 2011
- [17] Darmawan, M. F., Yusuf, S. M., Muhammad, A. R., and Haron, H., "Hybrid PSO-ANN For Sex Estimation Based on Length of Left Hand Bone", *IEEE Student Conference on Research and Development (SCOREd)*, Kuala, Lumpur, pp.478-483, 2015
- [18] Dileep, M. R., and Ajit, D., "Human Age and Gender Prediction Based on Neural Networks and Three Sigma Control Limits", *Applied Artificial Intelligence*, pp. 281-292, Vol 32, Iss 3, 2018
- [19] Gavin, H., and Ruff, C., "Sexual Dimorphism in Skeletal Brow Ridge and Chin Morphologies Determined Using a New Quantitative Method", *American Journal of Physical Anthropology*, pp.661-670, Vol 147, Iss 4, 2012
- [20] The German Online Portal for Statistics. [www.statista.com](http://www.statista.com). [Accessed: Oct 25<sup>th</sup> 2019].