IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.6, June 2022 E-ISSN: 2347-2693

An Improved Model For Baby Gender Guide Predictive System Using K-Nearest Neighbour Algorithm

Ezikwa Tenas God'swill[1*], Ezikwa Victoria Tenas[2]

[1]Department of Computer Science, Captain Elechi Amadi Polytechnic, Rumuola, Nigeria
[2]Department of Human Kinetics, Health and Safety Education, Ignatius Ajuru University of Education, Rumuolumeni, Nigeria

Corresponding Author: tenasandvick@gmail.com, Tel.: +2347065418012

DOI: https://doi.org/10.26438/ijcse/v10i6.915 | Available online at: www.ijcseonline.org

Received: 24/May/2022, Accepted: 12/Jun/2022, Published: 30/Jun/2022

Abstract- In almost every homes, having the desired gender of baby present could also foster the joy needed for the coexistence between couples in the family whereas in some instances, not having the desired gender of baby becomes the root cause of every other family problems. This research focuses on: "An Improved Model for Baby Gender Guide Predictive System using KNN classification algorithm". The model uses the trained dataset for prediction directly. The predictions were made by going through the trained dataset to obtain a new instance (x) for nearest neighbors and displaying the result of K instances. The new system was designed using object oriented analysis and design.methodology and was implemented using Hypertext Preprocessor (PHP) programming language and MySQL as the database software. The result of the new system indicates that the accuracy of the gender of babies predicted prior-to and within the first trimester of conception had a higher degree of accuracy of 92% which is superior to the sonographic system with an accuracy of 54%.

Keywords-Improved, Model, Computerized System, Effective, Baby, Gender Guide and Validation

I. INTRODUCTION

A baby gender guide predictive system is defined as a system that provide its users with tools capable of discovering and analyzing the outcome of a baby being a male or female prior or after completion of gestational period. From the user's point of view baby gender guide predictive system remains a black box that forecast an outcome whether a baby is a male or female. A baby gender guide predictive system focuses on data processing, filtering and outputting outcome; furthermore, it also cover aspects related to how users interact with it, including how to input data, how to define and evolve the user model, how to present the information to the users and how the users can manipulate that information [9]. Expectant couples become overwhelmingly excited, more curious and thereby become engaged in diverse of imaginations. Such thoughts or imaginations may include guesses of which of the couple or family member will the baby resemble, which complexion the baby will have and more importantly; what will be the gender of the expected baby because they are aware that in most instances, not having the desired gender of baby in a family results to problems such as: insecurity, acrimony, infidelity, abortion, loss of the womb, foetal deformity, polygamy, divorce and untimely death. The main motivating factor for embarking on this research study is the desire to improve and sustain the joy in every homes by minimizing or completely eliminate the list of problems associated with not having the desired gender of baby present in the family caused by the usage of inaccurate gender guide techniques. The aim of this study is to develop an improved model for Baby Gender Guide Predictive system using KNN algorithm whereas the specific objectives for the study include: design the new system using object oriented analysis and design methodology, Implementation of the TBGG predictive model using PHP programming language. The motivation for the study is to reduce the rate of divorce among couples due to their inability to get the desired gender of baby and foster the joy needed for the coexistence between couples in the family. This research work is very necessary because it will be immensely beneficial to: The General Public: it will help the general public especially most families in the society to uphold societal values by adhering to societal norms, educational health institutions: the result of the Baby Gender Guide Predictive system may serve as a literature that will help in providing better understanding on the positive and negative consequences of utilizing or not using an improved, accurate gestational gender guide system during childbearing; students of medical and para-medical studies: it can be used as a study guide in the following fields of human endeavours: Medicine and surgery, Nursing, Human Physiology, Human anatomy and Environmental health; Medical Health Facilities: Also, it can be used as a psychological and social therapy tools for building up the social and psychological wellbeing of couples who feel depressed as a result of not having the desired gender of baby in their family; researchers: it will serve as a spring board for researchers whose desire is to

further research on Baby Gender Guide Predictive. With the bid to ameliorate the existing biological baby gender determination method and the Chinese baby gender determination method, this study was embarked upon so as to combine the two methods biological baby gender determination method and the Chinese baby gender method by encapsulating them to develop an improved baby gender guide predictive system called Tenas Baby Gender Guide Predictive System so as to enhance the prediction of the gender of a baby without depending on the ultrasound gender guide system.

II. RELATED WORK

The term system is derived from the Greek word "Systema" which means an organized relationship among functioning units or components. A system that exist within the main system is referred to as a subsystem; the subsystems that make up the main system are designed to attain their individual objective or specific task first and in turn operate together so as to accomplish the overall aim of the main system. This elaborates how vital it is to consider the totality of the relationship a system has with its internal and external environment because a collection of components must work together to realize the aim of a system. [1] opined that a baby gender guide predictive system is designed and developed to forecast a baby's gender; but in most instances it does not settle the desired gender determination problem because it can only output its predicted result from several weeks after conception. The effects of the usage of inaccurate baby gender guide predictive system include: inaccuracy of information generated by the system thereby resulting to recurring incidence such as loss of the usefulness of the womb due to abortion of unwanted pregnancies, whereas on the other hand foetus that refuses to succumb to the pressure of being aborted are born with deformities [15]. The baby gender guide predictive system entails two main facets which are: biological principles for baby gender guide and non-biological systems for baby gender guide prediction. The related work for this study include the following:

[9] carried out a study on: "Accuracy of sonographic foetal gender determination: Predictions made by sonographers during routine obstetric ultrasound scan" The methods adopted include: A prospective, cross sectional study in a specialist private obstetric practice in the Illawarra, NSW. Although, the researchers were able to use obstetric ultrasound scan to determine foetal gender at a success rate of 100% accuracy after 14 weeks of conception; but the results varied depending on sonographer experience thereby resulting to a success decline of 54% for foetus younger than 12 weeks and the system was not able to predict the gender of the baby prior to conception.

[5**]** conducted a study on: "The ultrasound identification of foetal gender at the gestational age of 11-12 weeks". The cross-sectional methodology was adopted for the study and was conducted on 150 women in their 11th and 12th weeks of pregnancy in Hamadan in 2014. Although, the present study had a final success of foetal gender identification accuracy of about 91% in the 11th and 12th weeks of pregnancy and it is much higher when compared to other studies in the same category. But, it was not able to predict the gender of the baby prior to conception.

Furthermore, [13] proposed a robust iris gender-identification, the method adopted was based on a deep convolutional neural network. Although, they were able to identify a person's gender from an iris image but, its implementation was GPU specific and cannot be adopted for baby gender determination prior or during conception.

[17] Gender prediction by using Local Binary Pattern and K Nearest Neighbor and Discriminant Analysis classifications. K-Nearest Neighbor (KNN) was used for classifying the images into male or female, whereas and Discriminant Analysis (DA) methods was also used. Although they were able to have better performance by cropping parts of the images then feature extraction and classification methods applied on the face part of the images but were unable to determine the gender of a baby.

[18] carried out a research on Systematic Comparison of Age and Gender Prediction on IMU Sensor-Based Gait Traces. Traditional and deep learning methods were adopted. The study showed that it is feasible to predict age and gender with a reasonable accuracy on gait traces of just a few seconds and Further illustrates the need to put in place adequate measures in order to mitigate unintended information leakage by abusing sensors as an unanticipated side channel for sensitive information or private traits but was unable to predict the gender of a baby.

Finally, [13] researched on "A Study on Various Techniques Involved in Gender Prediction System" they adopted a comparative analysis research methodology for the study. Although, they were able to predict the gender of their subjects based on their handwriting with the aid of feature extraction and classification; but, they were unable to predict the gender of a baby prior conception because of their focus on using the result of the research for forensic study.

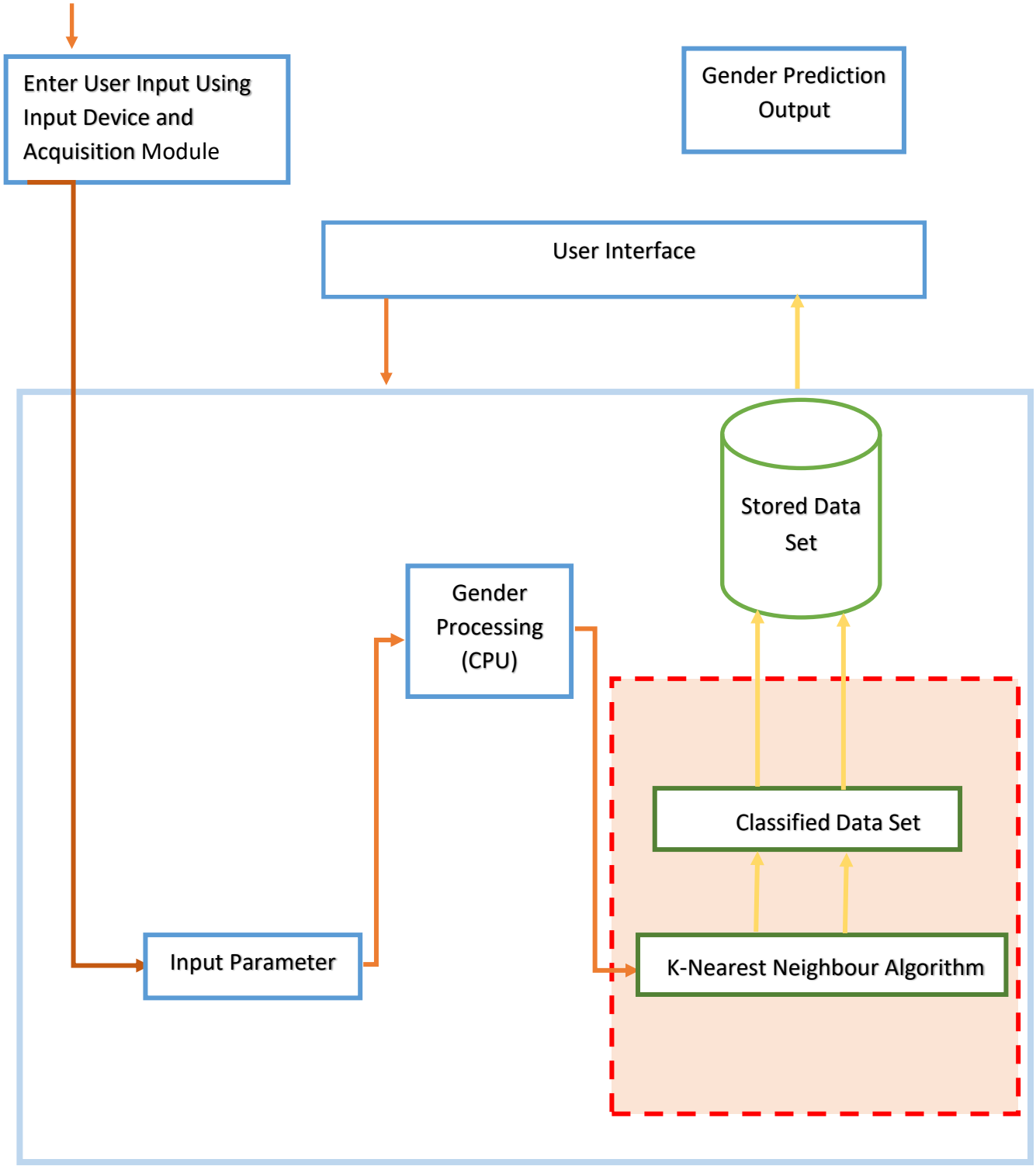
III. METHODOLOGY

In order to achieve the development of an effective baby gender guide predictive system, the Object-Oriented Analysis and Design (OOAD) methodology or paradigm was used because it is based on objects and classes. The OOAD methodology is suitable for this study because it is an internationally accepted software engineering methodology which is mainly used in most predictive system for identifying objects of a system to be designed, structure the software as a collection of separate objects that integrates both data structure and behavior, brings an abstraction of the real world based on objects including their interactions with other objects and identifying the relationship; because it has advantage over

the traditional approach in dealing with complexity and the fact that most languages are object oriented, it simplifies the software development and maintenance, it provides a conceptual foundation for assembling systems out of components using technological features such as: Class, Object, Polymorphism, Encapsulation, Abstraction and Inheritance which are the basis required for the development of the proposed system.

The new system will be structured to work with the standard software development procedure.

The architectural design of the proposed system shows the main component features of the new system and how its improvement on efficient baby gender guide prediction leveraged on the KNN algorithm.



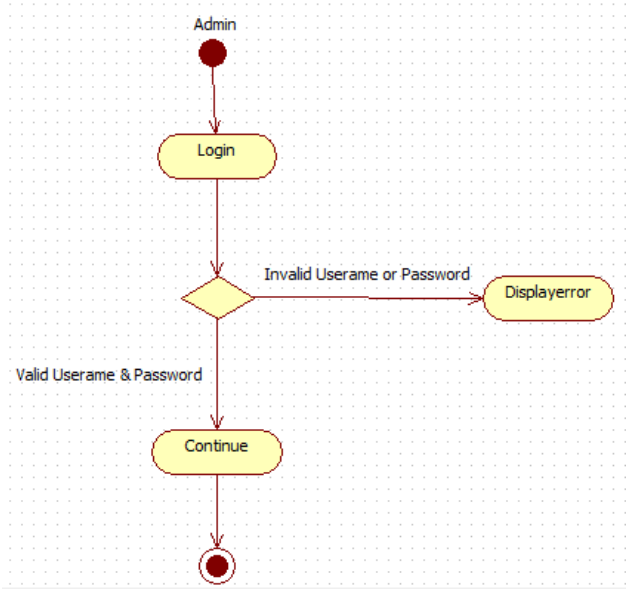


Figure 1: Architecture of the Proposed System

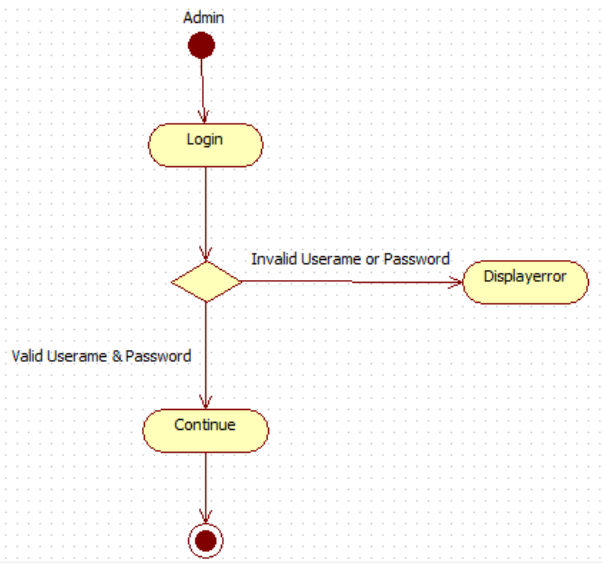


Figure 2: Activity Diagram of TBGG Predictive System

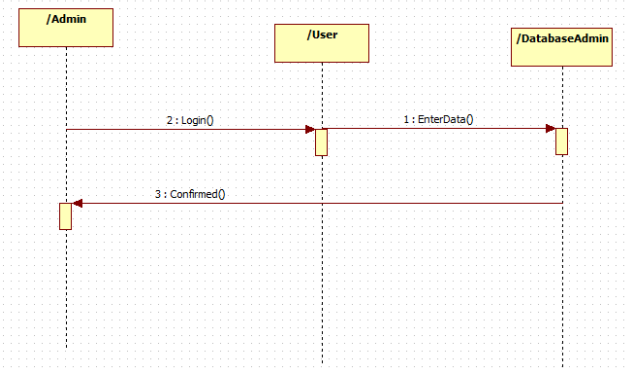


Figure 3: Sequence Diagram of TBGG Predictive System

K-Nearest Neighbors (KNN) Algorithm and Application

K-Nearest Neighbors (KNN) algorithm is a classification algorithm capable of storing the entire dataset such that no learning is require and the model uses the trained dataset for prediction directly. Predictions are made for a new instance (x) by searching through the entire training set for the K most similar instances (the neighbors) and summarizing the output variable for those K instances. To determine which of the K instances in the training dataset are most similar to a new input a distance measure is used. For real-valued input variables, the most popular distance measure one of these: Euclidean distance, hamming distance, Manhattan distance, Minkowski distance. The best distance metric is chosen by the researcher based on the dataset properties. Varying distance metrics and K values can be experimented if unsure of the given dataset so as to obtain the most accurate models.

When KNN is used for classification, the output can be calculated as the class with the highest frequency from the K-most similar instances. Each instance in essence votes for their class and the class with the most votes is taken as the prediction.

Class probabilities can be calculated as the normalized frequency of samples that belong to each class in the set of K most similar instances for a new data instance. For example, in a binary classification problem (class is 0 or 1):

p(class=0) = count(class=0) / (count(class=0)+count(class=1))

If you are using K and you have an even number of classes (e.g. 2) it is a good idea to choose a K value with an odd number to avoid a tie. And the inverse, use an even number for K when you have an odd number of classes.

Ties can be broken consistently by expanding K by 1 and looking at the class of the next most similar instance in the training dataset.

In this study, the KNN algorithm is used for classification of the gender of the baby into either male or female category and it is solely dependent on the mother's age and the month of conception, the output can be calculated as the class with the highest frequency from the K-most similar instances. Each instance in essence votes for their class and the class with the most votes is taken as the prediction.

K-Nearest Neighbour algorithm stores all available cases and classifies new cases based on a similarity measure, KNN can be used for both classification and regression problems. The algorithm uses 'feature similarity' to predict values of any new data points (e.g., distance functions). This means that the new point is assigned a value based on how closely it resembles the points in the training set. KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique.

Hamming Distance: It is used for categorical variables. If the value (x) and the value (y) are same, the distance D will be equal to 0. Otherwise D=1. Once the distance of a new observation from the points in our training set has

been measured, the next step is to pick the closest points. The number of points to be considered is defined by the value of k

$D_H =$ ---------- (1)

$x = y \; D = 0$ ---------- (2)

$x = y \; D = 1$ ---------- (3)

Where D_H is the harming distance

K, is the majority vote of its neighbours

X, is the first case of the categorical variable

Y, is the second case of the categorical variable

Table 1: Classification of a Case and Harming Distance

Gender		Distance
X	Y	
Male	Male	0
Male	Female	1

From table 1, a case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K-Nearest Neighbours measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor. It should also be noted that all three distance measures are only valid for continuous variables. In the instance of categorical variables such as gender, the Hamming distance must be used. It also brings up the Issue of standardization of the numerical variables between 0 and 1 when there is a mixture of numerical and categorical variables in the dataset. Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise but there is no guarantee. Cross-validation is another way to retrospectively determine a good K value by using an independent dataset to validate the K value. Historically, the optimal K for most datasets has been between 3 and 10.

IV. RESULTS AND DISCUSSION

This section discusses prediction of the baby's gender using the KNN algorithm. On inputting the mother's age and month of conception.

KNN-Prediction Using K=3 with Month of Conception = September and Mother's Age = 21 Years for Baby Gender Prediction

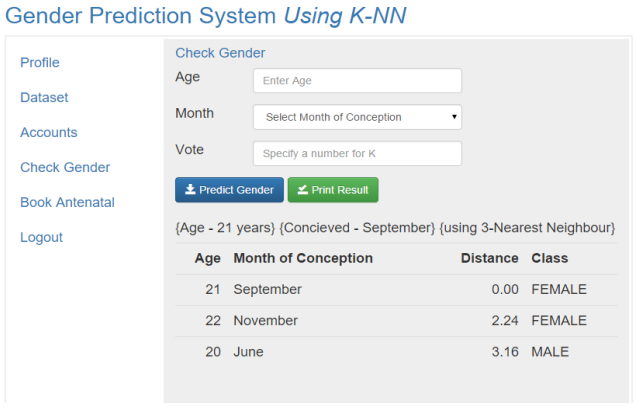


Figure 4a: Baby Gender Prediction Using K=3 Output Page

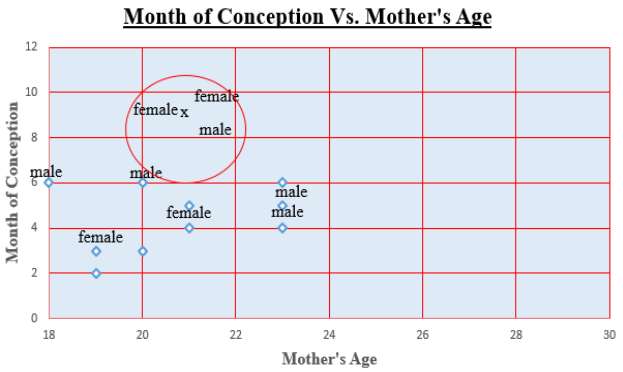


Figure 4b: Baby Gender Prediction Using K=3

Figure 4a shows the program output page whereas in figure 4b, shows a plotted graph of the month of conception versus mother's age, from the analysis it was found out that for a given odd number value of K = 3 with the month of conception provided as the 9th month and the mother's age is taken to be 21 years. The conclusive result from the above analysis using KNN algorithm shows that the proposed system predicted that the gender of the baby represented with the case (21, September) will result to a "female" gender because two out of the three nearest neighbours to the point labelled "x" were predicted to be data instances of babies with a gender of female thereby representing a percentage of female predicted gender of baby to be 66.67% which confirms that the result of the predicted gender of the baby will be a baby with a gender of "female".

KNN-Prediction Using K=5 with Month of Conception= September and Mother's Age = 21 Years for Baby Gender Prediction

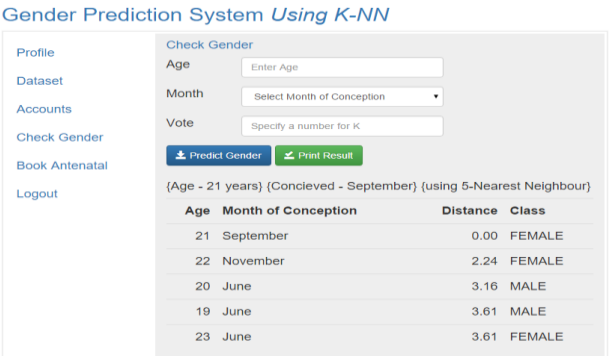


Figure 5a: Baby Gender Prediction Using K=5 Output Page

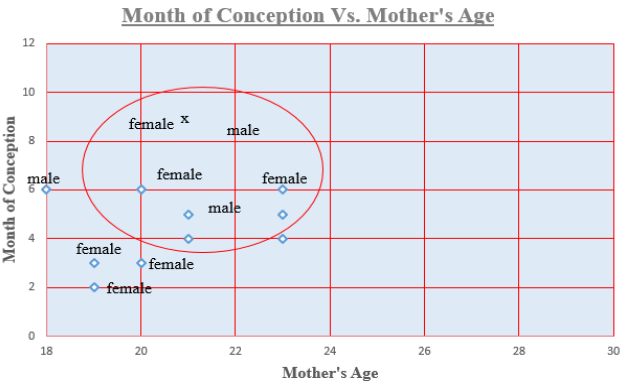


Figure 5b: Baby Gender Prediction Using K=5

From figure 5b above, depicts the graph plotted on the month of conception versus mother's age. If the vote (K) with a value of K=5 is used to measure the three other data instances of the class nearest to the point labelled "x" while the month of conception provided is the 9th month and the mother's age is 21 years. From the analysis above, it was drawn that using KNN algorithm, the system to a larger degree of accuracy predict that the gender of the baby represented with the case (21, September) will result to a baby with a gender of "female"; because, the three out of the five nearest neighbours to the point labelled "x" were predicted to be data instances of babies with a gender of female which represents a percentage ratio of male to female gender of baby to be 40.00% to 60.00% which is an affirmation that the result of the predicted gender of the baby will be a baby with a gender of "female".

KNN-Prediction Using K=5 with Month of Conception = February and Mother's Age = 23 Years for Baby Gender Prediction

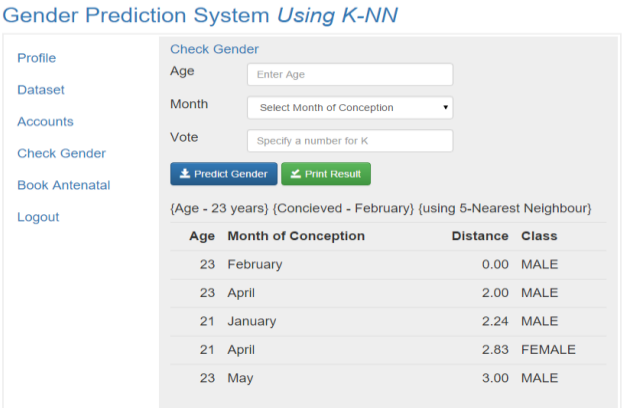


Figure 6a: Baby Gender Prediction Using K=5 Output Page

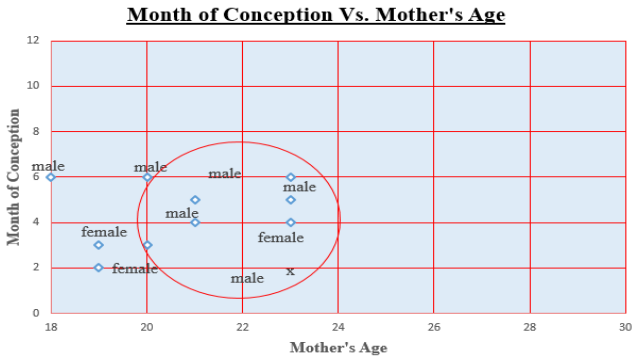


Figure 6b: Baby Gender Prediction Using K=5

Figure 6a shows the program output page whereas the graph in figure 6b shows the month of conception versus mother's age. The result of the analysis provides a better and clearer understanding on gender prediction using Tenas Baby Gender Guide Predictive System. The graph shows the month of conception versus mother's age. If the vote (K) with a value of K=5 is used to measure the three other data instances of the class nearest to the point labelled "x" while the month of conception provided is the 2nd month and the mother's age is 23 years. From the analysis above, it was drawn that using KNN algorithm, the system was able to a larger degree of accuracy predict that the gender of the baby represented with the case (23, February) will result to a baby with a gender of "male"; because, the three out of the five nearest neighbours to the point labelled "x" were predicted to be data instances of babies with a gender of male which represents a percentage ratio of male to female gender of baby to be 80.00% to 20.00% which is an affirmation that the result of the predicted gender of the baby will be a baby with a gender of "male".

V. DISCUSSION

The results obtained from the above analysis on the average gave 92% of accuracy in the prediction of the gender of a baby using the new system developed for this research as opposed to the 54% accuracy in the prediction of the gender of a baby obtained using the existing sonographic system. This was in consonance with the findings that the month and age of mother are very necessary for gender prediction while using both the new and existing baby gender guide predictive system [9].

VI. CONCLUSION AND FUTURE SCOPE

Gender prediction is a life critical process and promotes harmony in marriage especially in scenario where the desired genders of babies are abundant in the family, it was therefore concluded that although the new system developed will be able to predict the gender of a baby, it does include (or support) creating baby for couples which is the sole responsibility of the spouses. In future, other researchers should improve on the newly developed system by developing a new algorithm or formulating an hybrid algorithm to increase the percentage of accuracy.

REFERENCES

[1]. Adhatrao, K., Gaykar, A., Dhawan, A., Jaha, R. & Honrao, V. "Predicting students' performance using ID3 and C45 classification algorithms", *International Journal of Data Mining and Knowledge Management Process*, **Vol. 3, Issue 5, pp. 39-52, 2013.**

[2]. Antonio, M., Aida, V., David, I., Lucas, M. & Joan, B. "Ontology-based personalized recommendation of tourism and leisure activities", *Journal of Engineering Applications of Artificial Intelligence*, **Vol. 3, Issue 5, pp. 633-638, 2013.**

[3]. Callen, P. W. (2008). Ultrasonography in obstetrics and gynecology (5th Ed.). Saunders Elsevier publisher.

[4]. Chelli, D., Methni, A., Dimassi, K., Boudaya, F., Safar, E., Zouaoui, B., Chelli, H. & Chennoufi, M. B. "Foetal sex assignment by first trimester ultrasound: A Tunisian experience", *Journal of Population Health Metrics*, **Vol. 29, Issue 3, pp. 145–148, 2009.**

[5]. Eze, C. U., Ezugwu, F. O. & Agbo, J. A. "Sonographic determination of foetal gender in the second and third trimesters in a private hospital in Enugu", *International Journal of Ultrasonography*, **Vol. *16*, Issue 5, pp. 292–96, 2010.**

[6]. Farideh, G. "The ultrasound identification of foetal gender at the gestational age of 11–12 weeks", *International Journal of Family Medicine and Primary Health Care*, **Vol. 7, Issue 4, pp. 210-221, 2018.**

[7]. Jie, L., Dianshuang, W., Mingsong, M., Wei, W. & Guangquan, Z. "Recommender system application developments for gender

decision support system", *2nd International Conference on Software Eng. and Data Mining,* **pp. 47-51, 2015.**

[8]. Kumar, P. N. V. & Reddy, V. R. "A Survey on Recommender Systems and its applications", *International Journal of Innovative Research in Computer and Communication Engineering,* **Vol. 2, Issue 8, pp. 201-204, 2014.**

[9]. Manette, K., Karen, P. & Ian, G. "Accuracy of sonographic foetal gender determination: predictions made by sonographers during routine obstetric ultrasound scans", *Australian Journal of Ultrasound Medicine,* **Vol. 17, Issue 3, pp. 80-83, 2019.**

[10]. Mebarki, M., Kaidi, R., Azizi, A. & Basbaci, M., (2019) Comparative efficacy of two-dimensional mode and color Doppler sonography in predicting gender of the equine foetus, *Veterinary World*, ***12*****(2) 325-330, 2019.**

[11]. Mehrbakhsh, N., Othman, B. I., Norafida, I. & Rozana, Z. "A multi-criteria recommendation system using dimensionality reduction and Neuro-Fuzzy techniques", *International Journal of Basic and Applied Sciences,* **Vol. *19* Issue 11, pp. 317-320, 2015.**

[12]. Ning, L., Ping-Chia, T., Yu-An, C., and Homer, H. C. "Music recommendation based on artist novelty and similarity", *IEEE 16th International Workshop on Multimedia Signal Processing*, **1-6, 2014.**

[13]. Payal M., Abhishek G., Manoj K. G. "A study on various techniques involved in gender prediction system", *Bulgarian Journal of Cybernetics Sciences and Information Technologies,* **Vol. 19, Issue 2, pp. 11-14, 2019.**

[14]. Rashma, M., & Remesh, B. K. R. "Recommendation system: a big data application", *International Journal of Emerging Trends in Science and Technology Impact Factor,* **Vol. 3, Issue 9, pp. 39-46, 2016.**

[15]. Ricardo, S., Isabel, G., Ana, C., Diego, L., Pilar, P., Paola, M. & Juan L. "Ultrasound measurement learning of foetal sex during the first trimester: does the experience matter?", *Dove Press Journal Research and Reports in Focused Ultrasound.* **Vol. 15, Issue 4, pp. 201-205, 2015.**

[16]. Rudas, I. J. & Fodor Z. "Intelligent systems", *International Journal of Computers, Communications and Control,* **Vol. 13, Issue 2, pp. 132-138, 2011.**

[17]. Seda C. & Gökhan S. "Gender prediction by using local binary pattern, k nearest neighbor and discriminant analysis classifications. *International Journal of Advanced information Technology,* **Vol. 5, Issue 3, pp. 37-40, 2016.**

[18]. Tim V. H., Giuseppe G., Enrique A. R., Davy P., and Wouter J. "A Systematic Comparison of Age and Gender Prediction on IMU Sensor-Based Gait Traces", PMCID, **Vol. 19, Issue 13, pp. 294-295, 2019.**

