INTERNATIONAL JOUNAL OF: SOCIAL HEALTH AND MEDICAL RESEARCH

TIMBOU-AFRICA ACADEMIC PUBLICATIONS
INTERNATIONAL JOURNAL, MAY, 2023 EDITIONS
VOL. 13 NO. 3 ISSN: 2373-1340

WEB BASED RISK FACTORS ANALYSIS, DEATH PREDICTION AND PREVENTION IN SOME COMMON TYPES OF CANCERS USING CORRELATION BASED SELF ORGANIZING MAP (CB-SOM) MACHINE LEARNING MODEL.

[1]A.D. ADENIYI; [2]U. I. SULAIMAN; [3]M.T. AJIBADE; & [4]H. MORA

[1]Department of Mathematics and Computer Sciences, Faculty of Science, University of Medical Sciences, P.M.B. 536, Ondo City, Ondo-State, Nigeria. [2]Department of Computer Science, Faculty of Science, Nasarawa State University , Lafia, Nigeria. *[3,4]Department of Computer science, College of science and technology, Kaduna polytechnic, p.M.B.2021, T/Wada, Kaduna.*

ABSTRACT

Cancer also known as malignant tumors and neoplasm is a large group of diseases that can affect any part of the body by rapidly creating abnormal cells that grows beyond their usual boundaries and then spread to adjoining body parts, which most of the time resulted to death of the patient. This has

Introduction

Background of the Study

Cancer is considered to be the leading cause of death worldwide. According to World Health Organization (WHO) report, 2022, cancer account for nearly ten (10) million death globally, in 2020, nearly one in six deaths is caused by cancer globally WHO(2022). An estimated 9.6 million deaths in 2018 globally are as a result of different types of

been adjudged to be the leading cause of death globally, common among them are, the breast, lung, colon, rectum, postrate, skin and stomach cancers. This research work uses a unique toolset to exploit machine learning technique to analyse risk factor in this ailment in order to predict level of infection and likely death in some patients and to as well suggest preventive measures to healthy patients on-line and in real time. The Correlation Based Self Organizing Map (CB-SOM) machine learning technique has been trained to be used to this effect. The pre-ordered feature ranking filtering (PFRF) feature selection method was adopted to select the best features for the predictive system before applying the CB-SOM machine learning algorithm. To achieve these, the database of selected patients diagnostic report of some selected types of cancers was developed, this was analyzed using the propose model. The results were analyzed and presented. Performance, comparison of the proposed model was also carried out alongside with three (3) other machine learning techniques which includes the Naïve Bayesian, the traditional SOM and the K-NN algorithms. The experimental results show an excellent performance of the present system over the baseline methods studied. The present system is capable of producing a faster, accurate and efficient on line, real time predictions to the patients consistently.

Keywords: Prediction, Prevention, SOM, Pre-order, Cancers.

cancers in which the most common types include breast, lung, colon, prostrate, cervical, kidney and rectum cancers (Ferlay etal. 2020, WHO 2020).

This study presents the design and implementation of Web-based risk factors analysis, death prediction and prevention in some common types of cancers using Correlation Based Self-Organizing Map (SOM) machine learning algorithm. The study aims at analyzing risk factors in these diseases

in order to predict possibility of death in some highly risked patients and to suggest preventive measures to low risk patients. This will be achieved by creating a web-application that can be accessed by any individuals to determine their risk level in these diseases. The major contributions of this study are highlighted below:

First the study propose the development of a web based risk calculation and predictive system using Correlation Based Self-Organizing map (CB-SOM) machine learning algorithm. The novel algorithm has been proved to be computational efficient, accurate with low run time and capability to overcome scalability problem common to many machine learning algorithms while increasing execution speed.

Second, in order to make sure that best features are selected at the preprocessing stage of the predictive system before applying the proposed CB-SOM machine learning algorithm, the present system proposes the adoption of a unique feature selection algorithm called pre-ordered feature ranking filtering (PFRF) feature selection method. The said method has been proved to be effective accurate and faster in selecting the best attributes for any predictive and classification algorithms.

Third, the work also presents the construction of a specific risk calculator, death prediction and prevention system in four most common type of cancers namely Cervical cancer, Breast cancer, Lung cancer and Kidney cancer, using an experimental website developed with PHP, HTMLS, CSL3 and Java programming languages at the front end and MySQL database management system at the back end run on XAMP/Apache HTTP server as hosting server. The designed website will be used for data acquisition, model extraction, database creation for the predictive system. The present risk calculator, predictive and preventive website will be able to aggregate risk factor data from individual patient, build his/her personalized risk factors, analyse same and predict a risk level to the individual or presents a preventive measure to the individual as the case may be, online and in real-time basis. The propose application is an online web application that is

capable of allowing a physician or a patient to respond to a set of risk factor questionnaire, that are considered to have significant impact on the risk of these ailments, in order to obtain estimates of the risk level in the individual. This will significantly assist the patient and the physicians to determine the risk level for patient with such ailment and to assist government at all level to set priority for public health system in order to reduce mortality in these patients and the society at large.

Performance analysis of the present system will be done through performance comparison of the propose system with three (3) other machine learning algorithm which are Naïve Bayesian, the K-Nearest Neighbor (K-NN) and the Traditional SOM algorithms.

Finally, the results of the experiment will be presented and the proposed system will be implemented online and in real time.

Review of related work

The recent breakthrough in artificial intelligence technology has made machine learning a widely research field in recent time. A great deal of research has been done on predictive and recommendation system based on machine learning technique. This section deals with review of related work that are of important to this study with a view to explore the contribution of some scholars in the field of machine learning and to provide better understanding of the concept of online medical diagnosis and predictive system. Specifically, the review is organized according to the following sub topics.

A review of related machine learning technique

Hurwitz and Kirsch (2018) described Machine learning as a form of AI that enables a system to learn from data rather than through explicit programming. However, Machine learning uses a variety of algorithms that iteratively learn from data to improve, describe data, and predict outcomes.

A study of the work of different scholars in the field of machine learning and predictive field shows that there are lots of algorithms and methods for machine learning operation, some of which have been thoroughly compared and studied with the present system, some of the understudied techniques includes, the K-Nearest neighbor, ID3, random forest, self-organizing map(SOM), case based reasoning, the Naïve Bayesian etc. Bhosale and Ade(2014). Our reviewed evidence shows that most of these methods are marred with several challenges ranging from being computational complex, scalability challenges, being prone to noisy data to inaccuracy predictions. Adeniyi Wai and Yongquan(2015), Adeniyi Wai and Yongquan(2016). The present CB-SOM is designed to overcome these challenges by providing an accurate, faster, scalable and computation efficient methods of prediction consistently.

Overview of some predictive system.

A predictive medical diagnosis system deals with learning models to predict patients' health support system for individual patients and physicians in the diagnostic, therapeutic or patients monitoring operation. Different scholars in the field of machine learning and predictive system have come up with different methods these includes the Heart disease prediction by Soni, Ansari and Sharma(2011), Anibaras, Anupriya and Iyengar(2010). Prediction of cardiovascular risk factor through retina fudus photograph by Poplin et al., (2018), Adeniyi Wai and Yongquan(2018). Most of these system performed well using their data set but, the present system sets to offer a more accurate and effective predictive system than the existing ones studied.

A review of related feature selection system

Many dimensionally reduction techniques have been proposed by scholars in the field of machine learning which includes: the Gini index, correlation methods, wrapper methods and information gain. (Adeniyi, Ajoge and

Sulaiman, 2021, Gracia-Pedrajas and Haro Gracia, 2012, Boulle, 2019). Our review shows that none of these methods are perfect, but the present system adopted the PFRF method due to the fact that it has been proved to be more accurate, scalable and computationally efficient with the ability to handle noisy and high dimensionally data. Adeniyi, Ajoge and Sulaiman, (2021).

Methodology

This section describes the series of processes and methods adopted in realizing the objective of the present system.

Experimental Design

The present risk calculation and predictive system is trained on historical medical cases data extracted from the cancer unit of the Nigeria 44 Army reference hospital Kaduna, Kaduna State, Nigeria for a period of ten (10) years (2012 to 2021). The medical cases history of the extracted patients is kept anonymous and secured.

The study subjects consist of cohort of patient who are diagnosed of four (4) selected types of common cancers' i.e. Breast, cervical, Lung and Kidney cancers'. The risk level of these patients will be categorized based on these case records. Those dead will be categorized as deadly risk, those critically ill will be classified as "High risk" those with symptoms but not yet ill will be categorized as "low risk" while those diagnosed but no symptom will be categorized as "No risk". Based on this, a new patient can be predicted to be in any of those categories, while preventive recommendations will be made to those in "No risk" category. This will be done accordingly to similarities in the new patient's case history.

The pre-processing feature selection algorithm will be presented, follow by the classical algorithm for the predictive system i.e. the Correlation Based Self-organizing Map (CB-SOM) algorithm. The implementation will be done, the results analyzed and presented. Performance comparison will be done

alongside the baseline methods, the results summarized, conclusion will be drawn and recommendation will be made for further study. The overall architecture of the entire system is presented in figure 1.

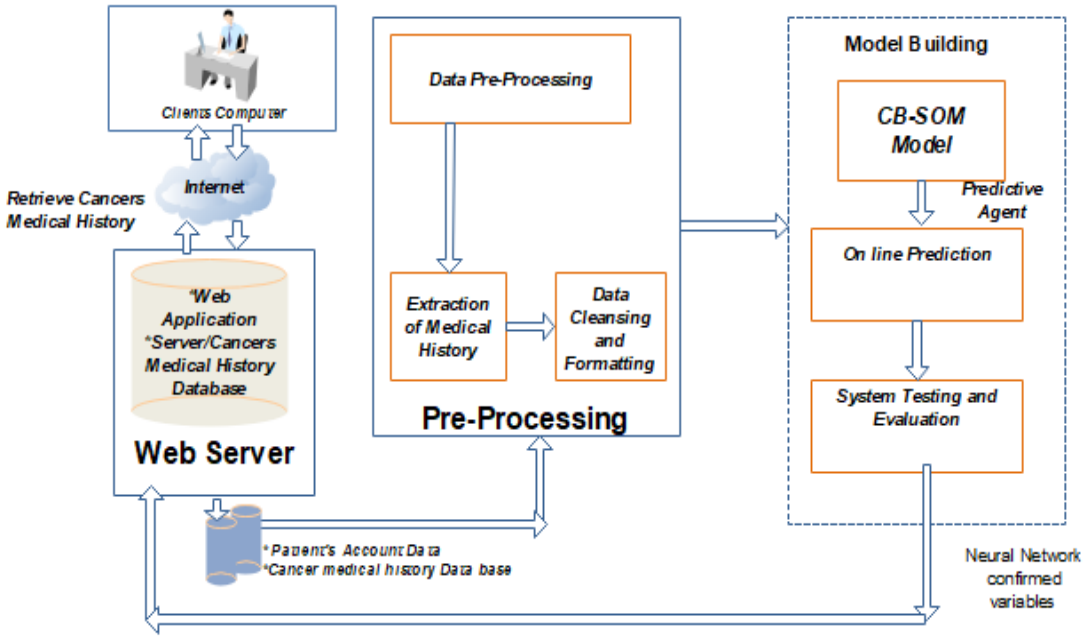

Figure1: The overall architecture of the entire system

Data Collection

In this work, a random selection of 15,948 medical case histories of anonymous cancer patients of 44 Nigerian Army reference hospital, Kaduna was made from the patient's database, the records consist of cohort of patients diagnosed of any of the four common cancers categories under study, for a period of ten years from 2012 to 2021.

The sample data are made up of 4,127 diagnosed of breast cancer, 3,752 diagnosed of kidney cancer, 2,850 diagnosed of cervical cancer and 2,016 of lung cancer.

The raw data extracted was cleansed and irrelevant/noisy data eliminated. After the data cleansing and noise elimination procedure, the adopted Pre-ordered Feature ranking Filtering (PFRF) feature selection algorithm was

used to select the best feature at the preprocessing stage of the predictive system, the patients database was developed, before applying the proposed Correlation Based Self Organizing Map (CB-SOM) machine learning algorithm.

The PFRF data Pre-processing feature selection Method

It is an established fact that most of the machines learning algorithms are marred with the challenges of intolerance to noisy and irrelevant features, Adeniyi, Ajoge and Sulaiman, (2021).

According to Adeniyi et al, 2018, feature selection is a basic dimensionality reduction technique adopted in machine learning and information retrieval system to select the most distinct feature in order to reduce high dimension data by selecting only useful attributes. A number of feature selection techniques have been studied for the purpose if this works some of which includes, the Gini index, correlation method, and information gam and wrapper method. It has been discovered that none of this methods are perfect. Adeniyi et.al., 2021.

However, the present system adopted the PFRF method, since it has been proved to be accurate, scalable, and computationally efficient and can handle high dimensionality and noisy data.

The PFRF uses the log scale to conduct term count in order to calculate the Term frequency (TF) measure for a given document d_i and attribute t_i, it takes the log of the sum of each attribute/ term using the expression.

$$\mathrm{FR}(t_i, d_i)^n = \log\left\{\sum_{n}^{i}\left((t_i, d_i)\right)\right\}\begin{cases}0, \text{If } n_{i,j} = 0\\ 1, \text{If } n_{i,j} > 0\end{cases} \qquad \text{Equation ... ... (1)}$$

J= 1,2,3,……n, i = 1,2,3, ……..n, n= 1,2,3,……m

n= Total number of documents, m= Total number of attributes.

Fr= Frequency for ranking the attributes.

The pre- ordered feature Ranking Filtering (PFRF) feature scaling method then discretized the log value into four levels using binning method as very high, high, moderate and low bins. The attributes are then ranked according to their bing values, so attribute with very high bin value only maybe selected for the present machine learning algorithm.
For more details on PFRF feature selection method see Adeniyi etal, 2021.

The Present Correlation Based Distance Measurement Method.

Evidence from available published literature studied and theoretical evidence shows that most of the existing distance measurement methods are marred with many challenges such as inaccuracy, prone to noisy data, computational complexity and scalability challenges to mention just a few.
The present correlation Distance based weight measurement technique is capable of overcoming these identified problems while providing accurate, simple, scalable, efficient, distance measurement for any machine learning technique.

The correlation relationship based weight measurement is used to measure relation and association between phenomena. The correlation relationship between two tuple X and Y can be expressed using the expression. Zaid (2015).

$$r(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})((y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad \text{Equation}\ldots\ldots\ldots 2$$

Where $\bar{x}$ is the mean of Tuple , and $\bar{y}$ is the mean of tuple y .
X= Training tuple
Y= test Tuple
Correlation can have a value of 1, 0 or -1
If the correlation is 1, then there is a perfect positive correlation between variable X and Y
If the correlation is 0, that means no correlation (the value of X and Y, don't seem linked at all)

If the correlation is -1, that means there is a perfect negative correlation. Zaid (2015).

The Proposed (CB-SOM) model

Neural network is a completely connected network of artificial neurons or nodes that made up a layered and feed forward node. It is used for the classification or estimation. The self-organizing Map SOM is a new form of neural network architecture, It is a type of unsupervised learning method used to discover some underlying structure of data. Zhang, Edwards and Harding (2007), Kohonon's SOM is a form of SOM widely believed to have been proposed and demonstrated by Tuevo kohonen in 1981. Zhang, Edwards and Harding (2007).

The idea is to represent a high dimensioned data in a low dimensional form without losing any essence of the data; it also serves to organize data on the basis of similarity through the arrangement of entities geometrically close to each other. The SOM method is useful in discovering association in a data set and clustering. Zhang, Edwards and Harding (2007).

In this work, effort was made to train neural network using extracted breast cancer patients' database. The training data is used to discover user group profile of a new patient and make a match to a specific patient risk level group in order to predict the risk level of the given patient online and in real time basis.

The working of CB- SOM model

In SOM, the model is used to group training tuples into cluster based on their common characteristics. These characteristics are organized into input vectors, these are obtained during data pre-processing and transformation. This is done by defining a binary factor for the entire risk factors questionnaire using binary vector to represent all questionnaire response string with one bit per questionnaire string.

This input vectors are denoted with an array of neurons as follows: Patient Y response to risk factors questioning can be represented as:
$Y= (q_1, q_2, q_3, q_4, q_n)$ where q_1 ……. q_n represent questionnaire;
$Y_1= \{0, 1, 0, 0, 1, 0, 1, 0, 1\}$, the same is applicable to a stored training tuple X.
$X_1= \{1, 0, 1, 0, 1, 1, 0, 1, 1\}$. The result of which can be represented in form a two dimensional output and referred to as a two dimensional Self-Organizing Map.

Figure 2 shows the illustration of the structure of a two dimension SOM.

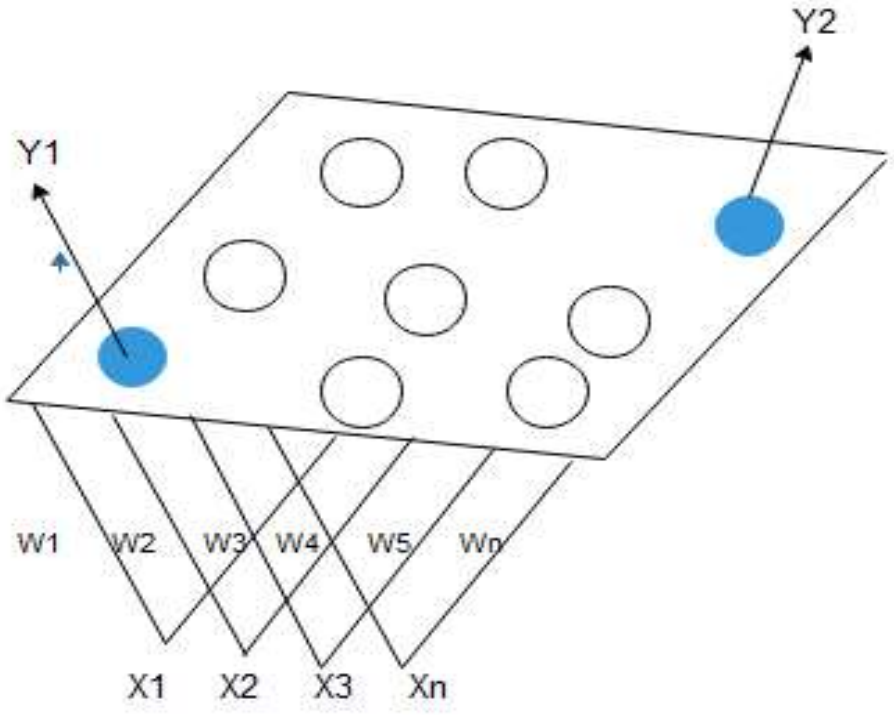

Figure 2: The Structure of a two dimension SOM.

At the training point, each tuple unit on the grid are compared with all the other tuples using Pearson correlation coefficient, if the correlation coefficient is 1, its weight and all other tuples in the clusters are adjusted for give a better match for the tuple, this process continues as training proceeds in order to form a two dimensional "map". When the learning

process is over, the given tuple is classified in the cluster as the training tuple. The weight update is given as shown in equation 2.

W_r (new) = W_r (old) +α[X_n- W_r (old)] Equation……. 3

= αX_n +(1-α)W_r (old), Equation …4

Where α is a term used to update the weight during training, it usually has a value ranges between 0 and 1.

W_r (old) = Weight before update for an output 1, W_r (new) = Weight after update for an output 1

X_n= input vector. , The present CB-SOM predictive machine learning algorithm is shown in Algorithm listing 1:

Algorithm listing

Algorithm listing 1: The present CB-SOM predictive machine learning algorithm

Let Y be an input patient's medical case history (q1, q2, q3, q4…… qn)

1. Function
2. Input: Training data set Xi (qi1, qi2, qi3, qi4……..…… q1n)

I= 1, 2, 3, ……..…… K

3. Let I = 1
4. Repeat until (perfect correlation)
5. Retrieve available patients medical case history(Yi) from the medical database
6. Compute the correlation between the given new case and the retrieved cases using the expression $r(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})((y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$
7. If the correlation between Xi and Yi = 1 (i.e. if r=1)
8. Then
9. Let Xi be a member of Yi

10. perform weight update using the expression W_r (new) = W_r (old) +α[X_n- W_r (old)], = αX_n +(1-α)W_r (old),
11. Go to 13; else
12. Increment I by 1
13. // Adapt the test case to fit in the new case to the correlated class
14. Evaluate the result
15. If (result work well) then,
16. Retain the new case history
17. Otherwise (discard the new case)
18. End if
19. End if
20. End function

System Evaluation and Results Analysis

In order to evaluate the performance of the present predictive using the proposed correlation based self or going map CB-SOM machine system, the results of the conducted experiment was applied.

System Implementation

In this work, an in-house software was developed using an experimental website with PhP, HTMLs, CSC3 and java programming languages at the front end and MySQL database management system at the back end running on XAMP / Apache HTTP server as hosting server. The developed software is capable of aggregating risk factor data from individual patient, physician, build his/her personalised risk factors, analyse and predict a risk level to the individual or recommend a preventive measure to the individual as the case may be online or in real time basis.

Figure 3a and 3b. show sample interfaces from the risk calculation and death prediction system that implement the proposed correlation distance based self-organizing map CB-SOM machine learning algorithm.

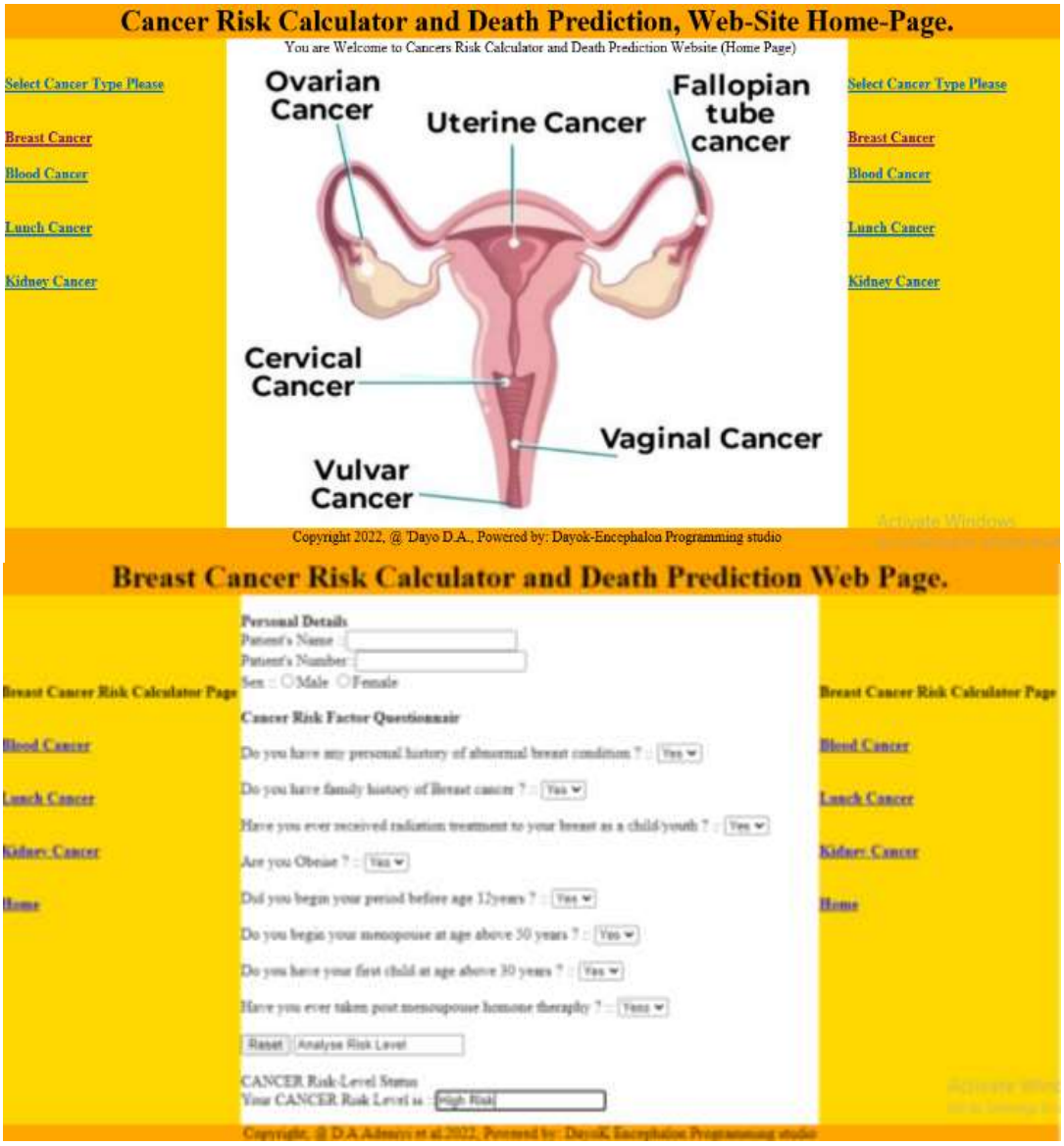

Figure 3a and 3b. Sample interface from the risk calculation and death prediction application.

System evaluation and results analysis

The performance evaluation of the present CB-SOM machine learning based risk calculator and death predictive algorithm was carried out to demonstrate the efficacy and capability of the system over some existing software which includes the traditional SOM, the K-NN and the Bayesian

techniques using the medical case history of four common types of cancers collected. The data was divided in to ten groups 70% of which was randomly selected as training set and the remaining 30% used as test set. The data was pre-processed using the adopted pre-ordered feature ranking filtering (PFRF) feature selection method; before applying the present CB-SOM Algorithm.

The present system was evaluated using F-Measure technique, we calculated the accuracy, True posture rate(TP), false posture(FP) rate, True negative TN, False Negative (FN), then the precision using TP, TN, FP and FN as defined in Adeniyi et al 2018. The F- Measure is the harmonic mean of precision and recall which is defined as

$$F1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

This is called F1 measure. This was computed using the present algorithm and the three baseline method which includes the traditional SOM, the Bayesian and the K-NM algorithms.

Figure 4a shows the F1 measure from our experimental results using cancer medical history data sets. While figure 4b is the speed of execution of the present algorithm and the baseline methods.

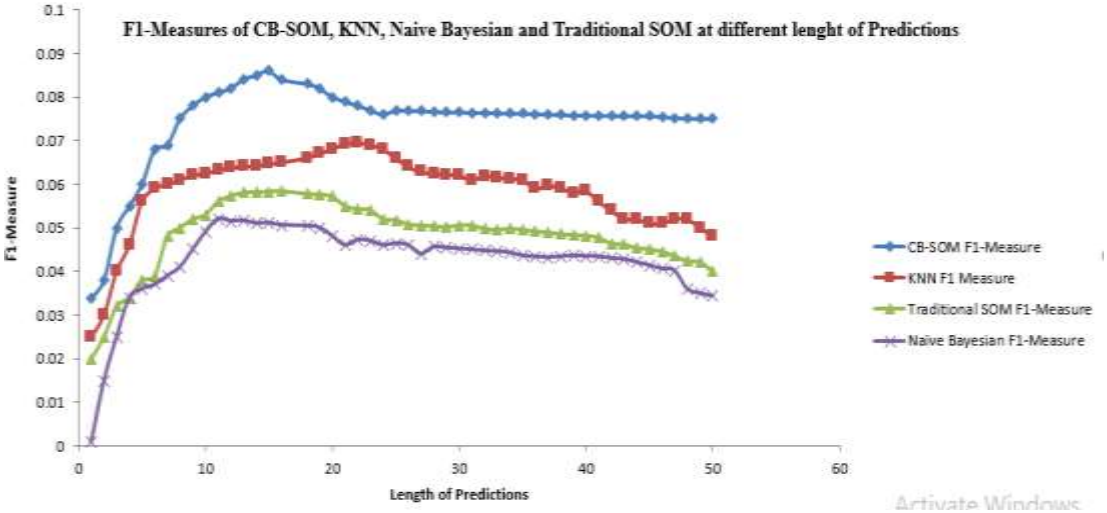

Figure 4a The F- measure from our experimental results using cancers medical history data sets collected.

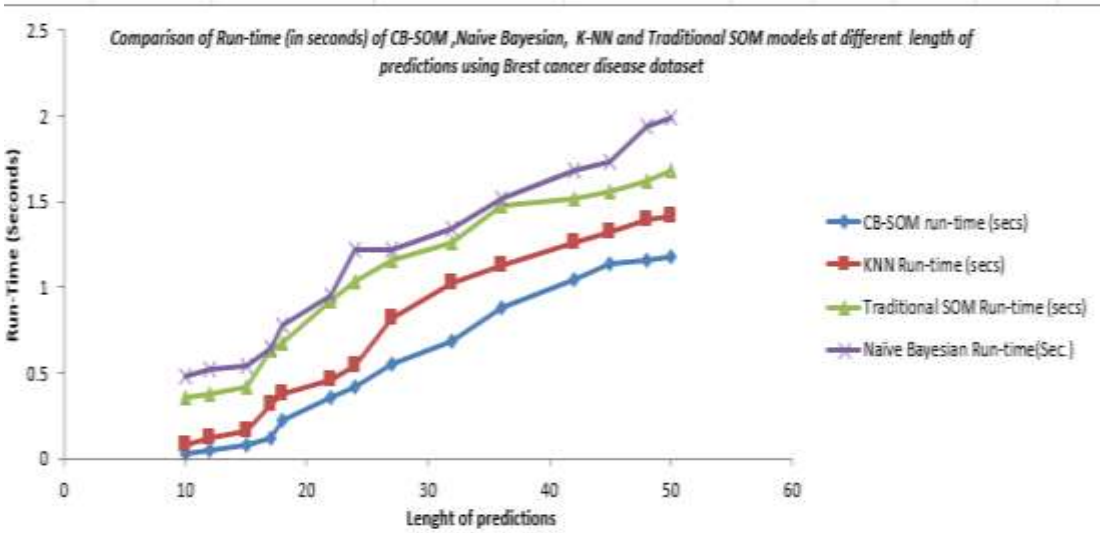

Figure 4b. The Execution speed of the CB-SOM model and the Baseline methods.

Discussion

The results of our experiment show an improved performance of the present CBSOM based risk factor analysis and death prediction in some types of cancers. The results show a high degree of accuracy over the baseline methods as shown in figure 4a, the result shows that the present system has over 80% degree of accuracy in comparison with the baseline techniques which hardly rise above 70%.

We further substantiate the improved performance of the CBSOM method over baseline technique by recording their run time of each of the model, in order to determine their speed of execution. The experimental results show the present CB-SOM runs faster than the other baseline method studied as shown in figure 4b, it produced an accurate, faster, efficient and consistent risk calculator and predications and recommendation to the user always online and in real time basis.

Summary, Conclusion and Recommendation

Summary of finding

This work is concerned with the development of a novel machine learning algorithm using the correlation distance based self-organizing map in order to analyze risk factors in some types of cancers ailment, predict death risk level and present recommendation to the user. The work also showcase the adoption of a unique feature selection techniques popularly referred to as Pre-order Feature Ranking Filtering (PFRF) algorithm in order to produce accurate, scalable and computational effective system capable of handling high dimensionally and noisy data before applying the present CB-SOM machine learning algorithm. To this effect an in house software was developed to implement the new model. The performance of the developed system was evaluated with three other baseline methods which are the Naïve Bayesian method, the K- nearest neighbor (K-NN) and the traditional SOM methods, the results show that the present system outperformed the baseline methods both in term of accuracy and speed

Conclusion and Recommendation

In this work, we proposed a novel risk factor analysis, death prediction and recommendation system using a novel machine learning algorithm known as Correlation Based Self Organising Map (CB-SOM) in order to aggregate risk factors in some common types of cancers analyse this data in order to predict death, risk level and present a preventive recommendation to the patients or physicians. The results of our experiment shows that the application of the present risk calculator system can produce accurate, faster and efficient recommendation/ prediction to the client consistently with low computational complexity.

We hereby recommend that the present work should be taken further by using a different types of similar data sets and further compare the performance of this system with other machine learning method in order to

further validate the efficacy of the present system in handling a problem of this nature in the future.

References

Adeniyi DA, Wei Z, Yongquan Y (2015) Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. J Appl Comput Inform. https ://doi.org/10.1016/j.aci.2014.10.001

Adeniyi DA, Wei Z, Yongquan Y (2016) Design and realization of online, real-time, web usage data mining and recommendation system using Bayesian classification method. Int J Comput Sci Eng Inf Technol Res (IJCSEITR). 6(3): 19–38 (**ISSN(P)): 2249–6831; ISSN(E): 2249–7943**).

Adeniyi DA, Wei Z, Yongquan Y (2018), Risk Factors Analysis and Death Prediction in Some Life-Threatening Ailments Using Chi-Square Case-Based Reasoning (χ2 CBR) Model. Interdisciplinary Sciences: Computational Life Sciences https://doi.org/10.1007/s12539-018-0283-6.

Adeniyi DA, Wei Z, Yang Y, (2016), Personalized news filtering and recommendation system using Chi-square statistics-based K-nearest neighbor (χ2SB-KNN) model. Enterp Inf Syst https :// doi.org/10.1080/17517 575.2016.12295 00.

Adeniyi D.A, Ajoge N.S.,Sulaiman U.I. (2021), Design and Realization of Pre-Ordered Feature Ranking Filtering (PFRF) Feature Selection Method for Machine Learning Algorithms. International Journal of Engineering and Technology Research Published by Cambridge Research and Publications IJETR ISSN-2329-7309 (Print) 82 Vol. 21 No.5 June, 2021. Pp. 82-100.

Anbarasi M.,, Anupriya, E., Iyengar,S.N. *Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm*, International Journal of Engineering Science and Technology Vol. 2(10), 2010, 5370-5376

Bhosale D, Ade R (2014). Feature selection based Classification using Naive Bayes, J48 and Support Vector Machine. Int J Comput Appl (0975–8887) 99 16.

Boullé,M.(2009), A parameter-free classificationmethod for large scale learning. J. Mach. Learn. Res. **10**, 1367–1385 (2009)

Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Pineros M. Global Cancer Observatory Cancer Today. Lyon: International Agency for Research on Cancer 2020(http://gco.iar.fr/today,accessed Febuary2021)

García-Pedrajas, N., Haro-García, A.(2012). Scaling up data mining algorithms: review and taxonomy. Prog Artif Intell (2012) 1:71–87 DOI 10.1007/s13748-011-0004-4.

Judith Hurwitz and Daniel Kirsch (2018), Machine Learning For Dummies, IBM Limited Edition, USA, John Wiley & Sons, Inc. 111 River St. Hoboken, NJ 07030-5774. ISBN: 978-1-119-45495-3 (pbk); ISBN: 978-1-119-45494-6 (ebk)

Mohamed Ahmed Zaid (2015), Correlation and Regression Analysis. The Statistical, Economic and Social Research and Training Centre for Islamic Countries (SESRIC) Kudüs Cad. No: 9, Diplomatik Site, 06450 Oran, Ankara – Turkey. www.sesric.org E-mail statistics@sesric.org.

Ryan Poplin,R, . Varadarajan A.V. Blumer K., Liu,Y., McConnell, M.V., Greg S. Corrado, G.S., Peng,L. and Dale R. Webster, D.R.(2018).Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning Nature Biomedical Engineering, Macmillan Publishers Limited, part of Springer Nature. https://doi.org/10.1038/s41551-018-0195-0.

Soni,J., Ansari, Z., Sharma, D. (2011).Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers. International Journal of Advanced Trends in Computer Science and Engineering 3(6).

World Health Organization(WHO) (2022) WHO Report on Cancer . www.who.int

World Health Organization, (2020), Assessing national capacity for the prevention and control of non communicable diseases report of 2019 global survey Geneva.

Xuejun Zhang, John Edwards, Jenny Harding (2007), Personalised online sales using web usage data mining. Computers in Industry 58 (2007) 772–782.