IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Predicting Cervical Cancer Prognosis using a Health Recommendation System.

Dr.R.G.Kumar[1] V Pradeepa[2] B Naga Bhushana[3] E Nandakumar[4] Y Pallavi[5]

[1]Associate Professor, [2345]UG Student, [12345]Department of Computer Science and Engineering,
[12345]Siddharth Institute of Engineering & Technology, Tirupathi, Andhra Pradesh, India,

ABSTRACT

The most common gynaecological malignancy that causes serious issues for women if left untreated is cervical cancer. This study exhibits various classification algorithms and demonstrates the benefit of feature selection methodologies for the most accurate cervical cancer disease prediction. The amount of patient digital system data in the healthcare industry is utilised to extract information and predict immune deficiency syndrome, which aids in patients' informed decision-making. To help patients understand the reports' material better, a health recommender system is employed. For feature selection in the proposed model, we employ wrapper techniques with the K Nearest Neighbor (KNN) classifier, and Multi Objective Algorithm (MOA) has been identified. is utilised in this system for feature selection as the most effective evolutionary algorithm to choose the key features with less complexity as compared to other conventional feature selection approaches. The dataset for the implementation and accuracy of the cervical cancer risk classification has been used as the assessment parameter.

Keywords*: K-Nearest Neighbor Algorithm, Multi Objective Algorithm, Feature Selection, Prediction, Health Recommender System.*

INTRODUCTION

Users receive services via recommender systems (RSs) based on an analysis of their historical data. They have demonstrated their effectiveness across the board. Recommender systems have made significant mistakes in industries like social networks and e-commerce. Health care is a crucial industry that necessitates the development of effective recommender systems to educate users by foretelling ailments based on their present lifestyle choices and a framework for an RS for the health sector. One of the illnesses that is considered to be the second greatest cause of mortality worldwide is cancer. Everywhere in the body, cancer can form and is a dangerous problem. One of the leading causes of death for women worldwide and the fourth most common malignancy is cervical cancer. The uncontrolled growth of abnormal cells in the cervix, a component of the female reproductive system, leads to cervical cancer. That is the main reason why women die from cancer in 42 different nations. Thus, detecting this illness early on is essential for saving the lives of women. The classification problem of predicting disease uses the two class values of prone to disease and not prone to disease. In classification problems, feature selection is a crucial preprocessing step.

The selection of features aims to balance two frequently at odds goals: reducing the number of features while also reducing the error rate of classification.

LITERATURE SURVEY

Software programmes known as recommender systems are able to deduce knowledge based on user data that contains information that is hidden. Technically, RS can be described as F: U I RN (1), where RN denotes the top-N items that the utility function F, which is responsible for producing RN, recommends for the user u U and for the item Iu I, in order to maximise the utility values for both the user u and the item Iu. The top-N recommendation problem is what this is called. RS is also known as a prediction problem, where it is forecasted whether or not a given item is intended for the user. The top-N recommendation and prediction problem has been addressed by a variety of health recommender systems (HRSs). A survey on HRS was suggested in [3] and examined various HRS implementation strategies and assessment techniques. A collaborative filtering method was developed to estimate cell line anti-cancer drug responses by combining cell line similarity networks and drug similarity networks on the basis that similar cell lines and similar drugs have similar responses. The drug response prediction was proposed as an RS problem in [4]. In [5], a Method is put out that uses drug and cell-line learning projections into a latent harmacogenomic space to predict cancer medication reactions to unseen celllines/patients.

A HRS with restricted boltzman machine and deep neural network has been presented in [6], which sheds light on how HRS may utilise big data analytics. A survey on HRS was suggested in [3] and examined various HRS implementation strategies and assessment techniques. A collaborative filtering method was developed to estimate cell line anti-cancer drug responses by combining cell line similarity networks and drug similarity networks on the basis that similar cell lines and similar drugs have similar responses. The drug response prediction was proposed as an RS problem in [4]. In [5], a Method is put out that uses drug and cell-line learning projections into a latent harmacogenomic space to predict cancer medication reactions to unseen celllines/patients. A HRS with restricted boltzman machine and deep neural network has been presented in [6], which sheds light on how HRS may utilise big data analytics. Single-objective optimization and multi-objective optimization are two types of optimization. Single objective optimization is the process of finding the best solution when there is only one objective function involved in the optimization issue. Multi-objective optimization is the process of locating one or more optimal solutions when the optimization issue calls for more than one objective function. The majority of search and optimization problems in the real world involve numerous priorities. There are multiple objective functions that need to be minimised or maximised in a multi-objective optimization problem. The two opposing criteria, such as common preferences and individual preferences, are taken into account by a multi-objective concept-based location recommender system [11]. In order to achieve these objectives, This method tested Foursquare and Gowalla datasets using a multi-objective evolutionary algorithm. In paper [12], a novel multi-objective-based menu recommendation is put forth that strikes the ideal balance between nutritional considerations, harmony, and coverage of pantry products.
Multi-objective Hydrologic Cycle Optimization (MOHCO), a multi-objective heuristic algorithm, is suggested in paper [13] to model water flow, infiltration, evaporation, and precipitation processes in nature in order to obtain the best possible collection of Pareto solutions. Using a dataset of 100K movie recommendations from Movie Lens, the proposed method is assessed, and the results show that MOHCO outperforms existing heuristic algorithms including MOEAD, NSGA-II, NSGA-III, and MOPSO.

A RS based on a multi-objective evolutionary algorithm [14] It has been suggested to create a set of non-dominated solutions with the twin goals of precision and diversity, where each solution represents a particular set of suggestions for the target user. With the use of the MovieLens and Jester datasets, a multi-objective evolutionary algorithm [15] based on decomposition that maximises the suggested systems' accuracy and variety has been tested. The process of choosing a subset of particular features or variables to be used in the creation of a model is known as feature selection in machine learning and statistics. That may be another factor that RSs consider when making suggestions that are appropriate based on a variety of item attributes. [16] has provided a thorough explanation of the various methods of feature selection. and parallel classification schemes for chronic illness forecasting. [17] proposes a content-based feature selection recommendation technique after investigating how feature selection affects content-based RS. [18] has built a system for the automated selection of content-based insightful features that is tested on RSs from diverse domains and irrespective of the type recommendation framework or kind of features. In order to preserve the victims' lives, predicting the likelihood of a cancer diagnosis is a need [19]. The best prediction algorithm has also been developed for breast cancer detection, together with a recommendation framework to generate a trustworthy suggestion for the prognosis of breast cancer [19]. Having reviewed the literature, It has been discovered that very little research has been done on HRS, and that the majority of those studies used conventional algorithms to make recommendations. Also, it has been noted that very little research on HRS employing multi-objective optimization and feature selection techniques has been conducted. In order to prevent the deaths of women from cervical cancer, this article has developed an RS that uses a feature selection approach based on MOGA [20] to forecast the prognosis of the disease's development.

CURRENT SYSTEM

In order to provide a trustworthy recommendation for the breast cancer prognosis and the prediction algorithm, a framework for recommendations is proposed. It has been recommended that a hybrid RS be used to give each patient a list of family doctor recommendations that takes into consideration the temporal complexity of their relationships. Using an akclique integrated deep learning classier, a diet recommender system is proposed that makes food product recommendations to patients based on their numerous health factors, such as blood pressure, cholesterol, sugar level, weight, etc.

A examination of the literature reveals that very little research has been done on HRS, and that the majority of those studies used conventional algorithms to make recommendations. Also, it has been noted that very little work has been done on HRS employing multi-objective optimization and feature selection techniques. In order to prevent the deaths of women from cervical cancer, this article has developed an RS that uses a feature selection approach based on MOGA [20] to forecast the prognosis of the disease's development.

PROPOSED SYSTEM

The k closest neighbour method is the suggested model. An efficient prognostic strategy must be developed that enables potential patients to receive earlier treatments before developing more severe diseases in order to reduce this enormous cost. Disease prevention is a laudable goal, and better prediction models enable high-risk individuals to receive preventative therapies in a timely manner, improving quality of life while sparing the country the financial burden of related treatment expenditures. Also, it is uncommon for illness risk prediction models to be verified using datasets that were not used in the training of the models.

The suggested Model is:

• K Nearest Neighbor algorithm. Accurate disease prediction is required to lower this enormous expense, and a successful prognostic system must be developed that enables potential patients to receive earlier therapies before developing more severe diseases.

• Disease prevention is a laudable goal, and better prediction models enable high-risk individuals to receive preventative treatments in a timely manner, improving quality of life while sparing the country the financial burden of related treatment expenditures.

• Datasets obtained separately from those used to train the models are rarely used to validate the development of disease risk prediction models.

OBJECTIVE

In order to prevent the deaths of women from cervical cancer, the goal of this research is to recommend a recommender system that uses feature selection to forecast the prognosis of the disease's development. Since that cervical cancer is the most curable type of cancer, risk assessment is unavoidably important in saving women's lives. This project aims to construct a predictive recommender system utilising a feature selection strategy based on the KNN Algorithm in order to inform women of the prognosis of cervical cancer according to their current life style activities.

TECHNIQUES TO BE USED FOR IMPLEMENTING THE PROJECT

To create the prediction models, we use the K-nearest neighbours technique to the dataset that was gathered. The KNN technique first balanced the dataset before it was used to train the model.

After fine-tuning the parameters and training, the average accuracy to accurately forecast the prevalence of disease as well as the average recall rate were determined. To train the dataset, we also used the KNN technique.

Project Architecture

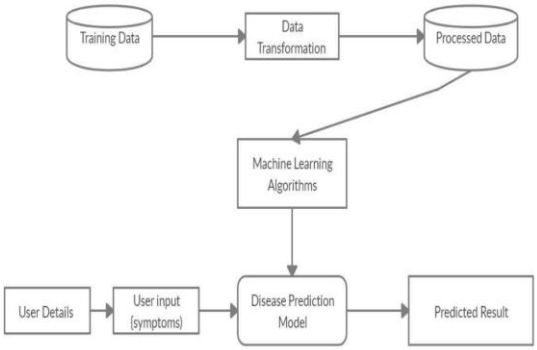

Fig 1: Architecture

Figure 1 depicts the structure for the health domain. Cancer is one of the diseases that is ranked as the second leading cause of death worldwide. Cancer can develop anywhere in the body and is a serious issue.

DESCRIPTION OF MODULES

Data Gathering and Preparation

The illness was the dataset used. Just the UCI Cleveland dataset, which is a mixture of four separate databases, was used. Although there are 76 properties in total in this database, all published experiments only mention employing a subset of 14 features.
For our investigation, we used the UCI Cleveland dataset that has previously been processed and is available on the Kaggle website. the full explanation of the 14 characteristics that were used in the suggested work.

Extraction of Features

In this, the original features are used to create a new set of features. The features must be transformed before being extracted. For feature extraction, Principal Component Analysis (PCA) is performed. Algorithms for linear transformations that are often utilised include principal component analysis.

K Nearest Neighbors, or KNN

When you are undertaking a pattern recognition task for classifying objects based on various attributes, the KNN method is helpful.
One of the simplest supervised machine learning techniques for classification is K-Nearest Neighbors. A data point is categorised depending on the classifications of its neighbours. It archives all cases in its database and groups fresh cases according to characteristics in common.
The best fit is provided by this algorithm for categorising our dataset.

Prediction Module

The administrator can introduce new symptoms in this module.
When a symptom occurs, For each condition with those specific symptoms, weights based on support and confidence variables are calculated and recorded in the database. Next, depending on the symptoms, the KNN algorithm was utilised to determine the real ailment.

Algorithm

U <- *User*
S <- *System*
BEGIN
Step 1: ***U*** *login ->* ***S***
Step 2: ***U*** *enter Uid, Psw ->* ***S***
Step 3: ***S*** *authenticate*
Step 4: ***S*** *give entry form ->* ***U***
Step 5: ***U*** *enter p name -> S*
Step 6: ***S*** *checks exist p name*
Step 7: if not exist then
goto step3

else
***S** perform prediction analysis*
End if
Step 9: checks is responsive positive/negative
*Step 10: **if** response positive **then***
updateScore
else
reward/penalizeScore
updateRating
give response->U
end if
END

RESULTS AND FINDINGS

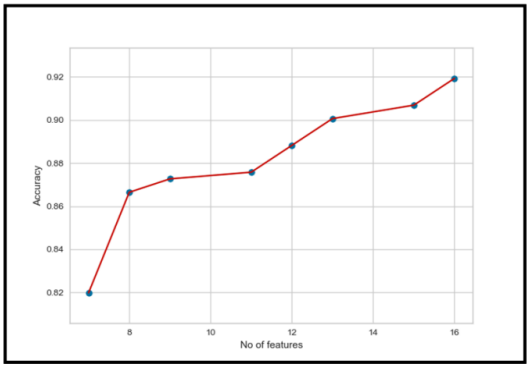

Fig 2: Comparison

We have classifiers and models for smaller datasets. The top-ranking model with the best accuracy ratings According to the implementation's results, the factors "Cancer," "HPV," "Dx," "Number of pregnancies," "CIN," "HIV," and "Age" are the most important factors for cervical cancer prognosis. Women need to be very conscious of these characteristics in order to lower the prognosis of cervical cancer. These characteristics' descriptions are displayed. The findings also suggest that models with GBM classifiers perform admirably and that models with Decision Tree classifiers exhibit the highest accuracy.

CONCLUSION & FUTURE WORK:

The use of machine learning algorithms to forecast disease is effectively deployed and produces results with higher prediction accuracy. The model not only distinguishes between cognitive impairment but also predicts the disease in the patient. The use of machine learning algorithms to forecast disease is effectively deployed and produces results with higher prediction accuracy. The model not only distinguishes between cognitive impairment but also predicts the disease in the patient.
Future research can combine several illness prediction methods with improved accuracy utilising machine learning algorithms. When they are combined, the disease may be more accurately anticipated at an earlier stage.

REFERENCES

[1] J. P. Burke, K. Williams, S. P. Gaskill, H. P. Hazuda, S. M. Haffner, and M. P. Stern, “Rapid rise in the incidence of type 2 diabetes from 1987 to 1996: results from the San Antonio Heart Study,” Archives of Internal Medicine, vol. 159, no. 13, pp. 1450–1456, 1999.

[2] C. D. Society, “Guidelines for the prevention and treatment of type 2 diabetes in China (2017 edition),” Chin J Diabetes, vol. 10, no. 1, pp. 4– 67, 2018.

[3] M.Wei,S.P.Gaskill,S.M.Haffner,andM.P.Stern,“Effectsofdiabetes and level of glycemia on all-cause and cardiovascular mortality: the San Antonio Heart Study,” Diabetes care, vol. 21, no. 7, pp. 1167–1172, 1998.

[4] A. J. Hanley, K. Williams, M. P. Stern, and S. M. Haffner, “Homeostasis model assessment of insulin resistance in relation to the incidence of cardiovascular disease: the San Antonio Heart Study,” Diabetes care, vol. 25, no. 7, pp. 1177–1184, 2002.

[5] M. Matsuda and R. A. DeFronzo, "Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp," Diabetes Care, vol. 22, pp. 1462–1470, Sep 1999.

[6] C. Lorenzo, K. Williams, and S. Haffner, "Insulin secretion based on the late oral glucose tolerance test period and incident diabetes: the San Antonio Heart Study," Diabetic Medicine, vol. 29, no. 8, pp. e151–e158, 2012.

[7] M. P. Stern, K. Williams, and S. M. Haffner, "Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test?," Annals of internal medicine, vol. 136, no. 8, pp. 575– 581, 2002.

[8] M. A. Abdul-Ghani, T. Abdul-Ghani, M. P. Stern, J. Karavic, T. Tuomi, I. Bo, R. A. DeFronzo, and L. Groop, "Two-step approach for the prediction of future type 2 diabetes risk," Diabetes Care, vol. 34, no. 9, pp. 2108–2112, 2011.

[9] M. A. Abdul-Ghani, K. Williams, R. A. DeFronzo, and M. Stern, "What is the best predictor of future type 2 diabetes?," Diabetes care, vol. 30, no. 6, pp. 1544–1548, 2007.