

Gastric Cancer Analysis In Data Mining Techniques

Barathiraja S, R. Karthiyayini

Master of Computer Application, University College of Engineering, Anna University, BIT Campus, Tiruchirapalli, Tamil Nadu, India

ABSTRACT

Gastric cancer is one of the most severe complex diseases with high morbidity and mortality in the world. The molecular mechanisms and risk factors for this disease are still not clear since the cancer heterogeneity caused by different genetic and environmental factors. With more and more expression data accumulated nowadays, a case of primary squamous cell carcinoma of the stomach is reported and the previously reported cases are reviewed. We can perform integrative analysis for these data to understand the complexity of gastric cancer and to identify consensus players for the heterogeneous cancer. In the present work, we screened the published cell expression data and analyzed them with integrative tool combined with pathway and cell ontology enrichment investigation. Considering not only the stage of disease, but also patients' age and comorbidities. Experimental validation is proposed to further confirm this finding.

Keywords: Data Mining, Integrative Analysis, Bootstrap Function, R Graphics,

I. INTRODUCTION

The Gastric cancer (GC) is currently the most common one of the most severe cancers in the world with high incidence and low survival rate. Although gastric cancer is showing a decrease in its incidence since the beginning of the 20th century, it still represents the fourth most common type of cancer, and the second leading cause of cancer related death worldwide (1). According to the global cancer statistics report in 2012, GC has been the than seven hundred thousand deaths each year [1]. Usually, the number of GC patients in men is twice more than that in women and Eastern Asia, especially Korea, Japan, and China, has the highest incidence rate. Although relevant reports revealed that the age-standardized incidence rate of gastric cancer is decreasing in Japan and Korea in last few years [2, 3], the number of new cases is still increasing due to the aging of the population. The pathogenesis of gastric cancer is very complex and remains unclear. Recent basic studies mainly focus on three main factors: environmental factors, *Helicobacter pylori* (*H. pylori*) infection, and gene expression deregulation [4, 5].

The method we used here was named INMEX (integrative meta-analysis of expression (program [17])). Data procession and screening were performed in order to make sure all the datasets we uploaded into the program were in a consistent format a primary gastric epidermis cancer by its location in the distal stomach though the pathologist included the possibility of esophageal origin in his report.

R is a wonderful programming language for statistics and data management, used widely in industry, business, government, medicine and so on. And it's free, an open source product. The S language, of which R is essentially an open source version, won the ACM Software System Award in 1998. R syntax is similar to those of C, Python, PERL, etc. (don't know these terms) object-oriented and has a functional programming philosophy. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team, of which Chambers is a member. R is named partly after the first names of the first two R authors and partly as a play on the name of S. The project was conceived in 1992, with an initial version released in 1994 and a stable beta version in 2000. R and its

libraries implement a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others. R is easily extensible through functions and extensions, and the R community is noted for its active contributions in terms of packages. Many of R's standard functions are written in R itself, which makes it easy for users to follow the algorithmic choices made. For computationally intensive tasks, c, c++, and FORTRAN code can be linked and called at run time. Advanced users can write C, C++, Java .NET or Python code to manipulate R objects directly. R is highly extensible through the use of user-submitted packages for specific functions or specific areas of study. Due to its S heritage, R has stronger object-oriented programming facilities than most statistical computing languages. Extending R is also eased by its lexical scoping rules.

Another strength of R is static graphics, which can produce publication-quality graphs, including mathematical symbols. Dynamic and interactive graphics are available through additional packages. R has Rd, its own LaTeX-like documentation format, which is used to supply comprehensive documentation, both on-line in a number of formats and in hard copy. The capabilities of R are extended through user-created packages, which allow specialized statistical techniques, graphical devices import/export capabilities, reporting tools (knitr, Sweave), etc. These packages are developed primarily in R, and sometimes in Java, C, C++, A core set of packages is included with the installation of R, with more than 7,801 additional packages (as of January 2016) available at the Comprehensive R Archive Network (CRAN) Bioconductor, Omegahat, GitHub, and other repositories. The "Task Views" page (subject list) on the CRAN website lists a wide range of tasks (in fields such as Finance, Genetics, High Performance Computing, Machine Learning, Medical Imaging, Social Sciences and Spatial Statistics) to which R has been applied and for which packages are available. R has also been identified by the FDA as suitable for interpreting data from clinical research.

Other R package resources include Crantastic, a community site for rating and reviewing all CRAN packages, and R-Forge, a central platform for the collaborative development of R packages, R-related software, and projects. R-Forge also hosts many

unpublished beta packages, and development versions of CRAN packages. The Bioconductor project provides R packages for the analysis of genomic data, such as Affymetrix and cDNA microarray object-oriented data-handling and analysis tools, and has started to provide tools for analysis of data from next-generation high-throughput sequencing methods.

II. METHODS AND MATERIAL

Interfaces:

- ✓ Architect – cross-platform open source IDE for data science based on Eclipse and StatET
- ✓ DataJoy – Online R Editor focused on beginners to data science and collaboration.
- ✓ Deducer – GUI for menu-driven data analysis (similar to SPSS/JMP/Minitab).
- ✓ Java GUI for R – cross-platform stand-alone R terminal and editor based on Java (also known as JGR).
- ✓ Number Analytics - GUI for R based business analytics (similar to SPSS) working on the cloud.
- ✓ Rattle GUI – cross-platform GUI based on RGtk2 and specifically designed for data mining.
- ✓ R Commander – cross-platform menu-driven GUI based on tcltk (several plug-ins to Rcmdr are also available).
- ✓ Revolution R Productivity Environment (RPE) – Revolution Analytics-provided Visual Studio-based IDE, and has plans for web based point and click interface.
- ✓ RGUI – comes with the pre-compiled version of R for Microsoft Windows.
- ✓ RKWard – extensible GUI and IDE for R.
- ✓ RStudio – cross-platform open source IDE (which can also be run on a remote Linux server).

Commercial support for R:

While R is an open source project supported by the community developing it, some companies strive to provide commercial support and/or extensions for their customers. In this section, some examples of those companies are mentioned.

In 2007, Revolution Analytics was founded to provide commercial support for Revolution R, its distribution of R, which also includes components developed by the

company. Major additional components include: ParallelR, the R Productivity Environment IDE, RevoScaleR (for big data analysis), RevoDeploy R, web services framework, and the ability for reading and writing data in the SAS file format-They also offer a distribution of R designed to comply with established IQ/OQ/PQ criteria which enables clients in the pharmaceutical sector to validate their installation of R Evolution R. In 2015, Microsoft Corporation completed the acquisition of Revolution Analytics.

For organizations in highly regulated sectors requiring a validated version of R, has developed the Valid R product which fully complies with the Food and Drug Administration guidelines for Software verification and validation-They also offer to validate additional packages if the customer demands it and validate customer's self-written packages.

III. LITERATURE SURVEY

Gastric Cancer Associated Genes Identified by an Integrative Analysis of Gene Expression Data Bing Jiang, 1 Shuwen Li, 2 Zhi Jiang, 3 and Ping Shao1 (*Department of Spleen and Stomach Diseases, Hospital of Traditional Chinese Medicine, Yixing, Jiangsu 214200, China*)

The pipeline of this whole analysis in the present study is shown in Figure 1. We first extract the microarray gene expression data from the GEO (Gene Expression Omnibus) database, then we integrate analyzed the expression data with a meta-analysis tool INMEX (integrative meta-analysis of expression data), and then we further screen and validate the meta-analysis results with literature analysis and bioinformatics functional analysis.

The Surgical Treatment of Gastric Adenocarcinoma a Literature Review

Thiago de Almeida Furtado¹, Lélis Sanglard Oliveira², Marília Ribeiro Lima Gramiscelli Costa², Luiza Ohasi de Figureiredo³, Diego Paim Carvalho Garcia⁴ and Luiz Ronaldo Alberti^{5*} (⁵Department of Surgery, Brazil)

We evaluated recent literature (from 2005 up to June 2016) on PubMed Central with combination of following MESH terms. Elderly and gastric cancer and

lymphadenectomy, elderly and gastric cancer and complications, high risk patients and gastric cancer.

Table 1 reports data from some of the considered comorbidities with aging (8), underlining the frailty of this population. Although in last year's literature reports many studies concerning oncologic gastric surgery in elderly patients, there is still no homogeneity in the definition of this subgroup, and an established cut-off age seems far to be determined: in facts, if some authors defined as elderly patients with more than 65 years (14), other authors reported thresholds of 70, 75 or 80 years (3, 11, 15-17), so that it is not easy to compare results from different studies.

IV. EXISTING SYSTEM

Gastric (including gastro esophageal junction) cancer is the third leading cause of cancer-related death in the world. Has evolved in recent years. Although systemic chemotherapy is still the mainstay treatment of metastatic disease. The introduction of agents targeting human epidermal growth factor receptor 2 and vascular endothelial growth factor. Several randomized control trials and meta-analyses have demonstrated the positive effects of adjuvant chemotherapy in patients with advanced gastric cancer, in terms of cancer recurrence and patient survival. The preliminary yet encouraging clinical efficacy observed with immune checkpoint inhibitors, e.g., anti-programmed cell death protein 1/programmed death-ligand 1, will further shape the treatment landscape for gastric cancer. Molecular characterization of patients will play a critical role in developing new agents, as well as in implementing new treatment options for this disease

V. PROPOSED SYSTEM

The proposed system make use of an analysis of Gastric cancer dataset. Its relatively simple and easy understand away. Use data mining algorithm supposed to analyses for r environment. The relevant Dataset is represent to hold the 2-to-5 year's data. When the Dataset is analyzed for using r algorithm. There are several algorithm used to support in r tool. And to define produce the output of a result is represents the graphical form. If the benefit of this project is similar to yearly no of patient remissible in the record set and the no of patients takes a higher level of treatments when the data's are maintained. And the final conclusion of the result is produced to the system.

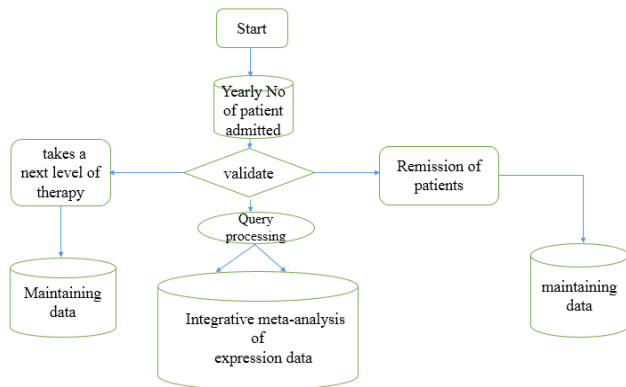


Figure 1. The architecture of the proposed system

SYSTEM SPECIFICATION

Hardware specification:

- SYSTEM: Pentium IV2.4 GHZ
- HARD DISK: 40GB
- MONITOR: 15VGA Color
- RAM: 1GB
- KEYBOARD: 110 keys enhanced

Software specification:

- OPERATING SYSTEM: Windows XP and above
- FRONT END: R Tool
- BACK END: Excel

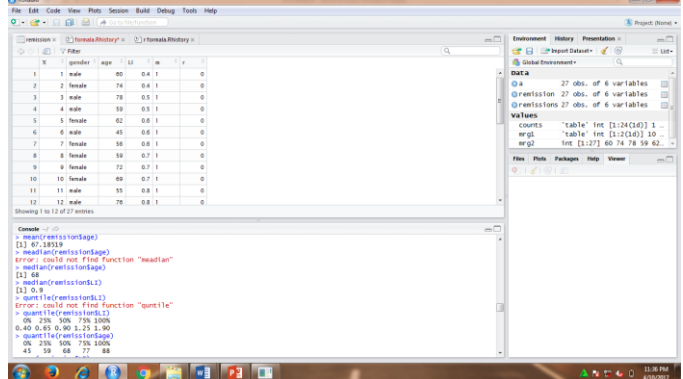
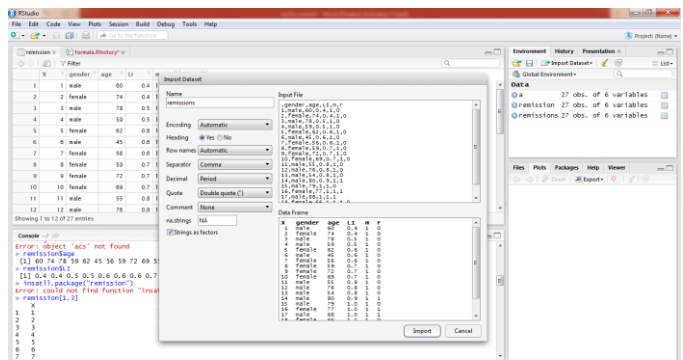
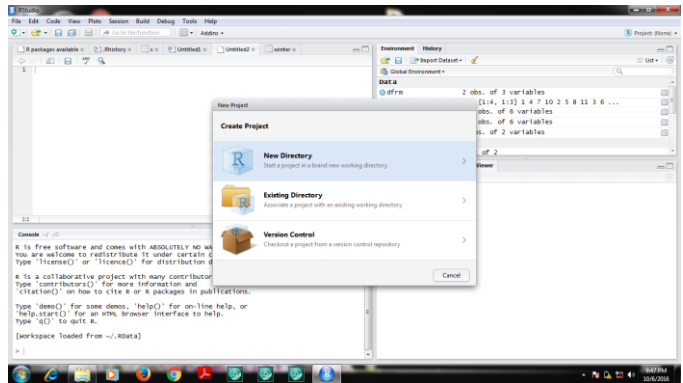
Implementation

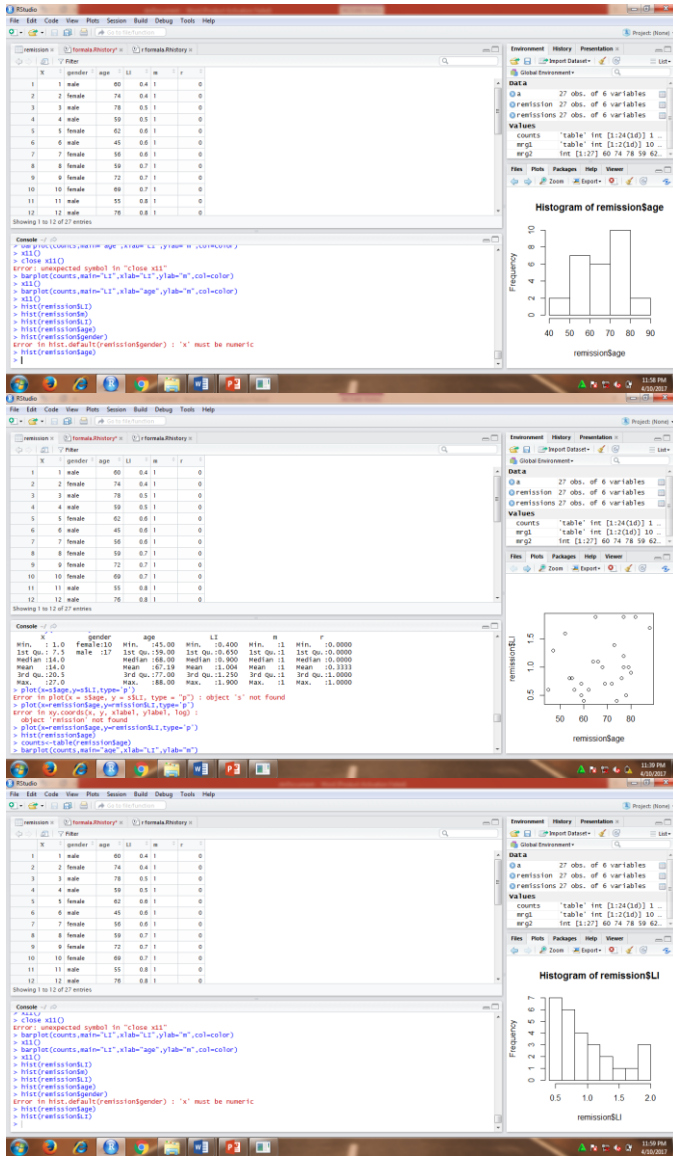
The implementation of the application is done in the following manner using XL for dataset, spring and Hibernate (r) for middleware and Angular numeric number for front end view layer. This modules we represent the User and patient details, User and patient details module maintains the details of the users. It also maintains the details of the all the patients of the hospital along with cancer patient details. This module also provides a dependent relationship with the patients and the related surgical treatments or projects based on their e which is responsible in maintaining the details. Project details module is the module which is responsible in maintaining the project details along with the activities to be completed for a project. The implementation of the project and model is, DAO layer (Interface and implementation), the method interface with query processing and integrative analysis for the gastric cancer dataset. The application following the Model-View-Controller architecture. (query processing)classification of the none-cancer patients

and cancer patients. The DAO layer has the interface and its implementation to analysis the dataset.to find the results is provided for graphical representation we screened the system.

VI. RESULTS

The Dashboard page is used to display the overall estimate the patient records based on the completion of projects and courses and top trending courses and projects based on the enrollment of the cancer of the department.





VII. CONCLUSION

To summarize, our research provides novel angels in pathogenesis of gastric cancer. We identified consistently DE genes in gastric cancer through INMEX meta-analysis tools. potentially Low cost and provide us low resolution data but that data is enough to provide warning to public. We identify An accurate preoperative evaluation of patients and an optimization of general conditions, together with an accurate staging of disease, are obviously necessary to tailor on patients the best surgical option, in order to avoid the risk of under-treatment in elderly patients who might benefit from a more aggressive approach. Which have never been reported to be associated with GC before. Although many improvements have been made in the early diagnosis and surgical treatment of gastric cancer. Recent studies have shown good efficacy of the procedure and demonstrate reduction of complications

and survival rates maintained or improved. Further experimental validation should be conducted in order to understand the mechanism of these genes on gastric cancer.

VIII. REFERENCES

- [1]. A. Tvarijonaviciute, J. J. Ceron, C. de Torre et al., “Obese dogs with and without obesity-related metabolic dysfunction— a proteomic approach,” BMC Veterinary Research, vol. 12, no. 1, article 211, 2016
- [2]. Ferro A, Peleteiro B, Malvezzi M, et al. Worldwide trends in gastric cancer mortality (1980-2011), with predictions to 2015, and incidence by subtype. Eur J Cancer 2014;50:1330-44.
- [3]. Boswell, J. T., and Helwig, E. B.: Squamous cell carcinoma and adenoacanthoma of the stomach. Cancer 18:181-192, 1965.
- [4]. GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) Group, Paoletti X, Oba K, Burzykowski T, Michiels S, Ohashi Y, et al. Benefit of adjuvant chemotherapy for resectable gastric cancer: a meta-analysis. JAMA 2010;303:1729-1737