Review on Pharmaceutical Industries Production of Medicines for Lung Cancer Diseases Prevention and Side Effects DataSets View and Analysis with Orange Software Visualization Techniques

V. Srinivasan [1] & S. Soumya [2]

[1] Research Scholar, Institute of Computer Science and Information Sciences, Srinivas University, Mangalore, Karnataka, India.

ORCID ID: 0009-0001-0923-0414; E-Mail: victorsrics@gmail.com

[2] Assistant Professor, Institute of Computer Science and Information Sciences, Srinivas University, Mangalore, Karnataka, India.

ORCID ID: 0000-0002-5431-1977; E-Mail: pksoumyaa@gmail.com

Area/Section: Computer Science
Type of the Paper: Review Paper
Type of Review: Peer Reviewed as per |C|O|P|E| guidance.
Indexed in: OpenAIRE.
DOI: https://doi.org/10.5281/zenodo.14769203
Google Scholar Citation: IJMTS

How to Cite this Paper:
V. Srinivasan & S. Soumya (2025). Review on Pharmaceutical Industries Production of Medicines for Lung Cancer Diseases Prevention and Side Effects DataSets View and Analysis with Orange Software Visualization Techniques. *International Journal of Management, Technology, and Social Sciences (IJMTS), 10*(1), 17-44. DOI: https://doi.org/10.5281/zenodo.14769203

International Journal of Management, Technology, and Social Sciences (IJMTS)
A Refereed International Journal of Srinivas University, India.

CrossRef DOI: https://doi.org/10.47992/IJMTS.2581.6012.0370

Received on: 16/11/2024
Published on: 30/01/2025

Review on Pharmaceutical Industries Production of Medicines for Lung Cancer Diseases Prevention and Side Effects DataSets View and Analysis with Orange Software Visualization Techniques

V. Srinivasan [1] & S. Soumya [2]

[1] Research Scholar, Institute of Computer Science and Information Sciences, Srinivas University, Mangalore, Karnataka, India.

ORCID ID: 0009-0001-0923-0414; E-Mail : victorsrics@gmail.com

[2] Assistant Professor, Institute of Computer Science and Information Sciences, Srinivas University, Mangalore, Karnataka, India.

ORCID ID: 0000-0002-5431-1977; E-Mail : pksoumyaa@gmail.com

ABSTRACT

Purpose: *The purpose of this research is to explore how Orange, a powerful information extraction and predictive modeling software, can be applied in the pharmaceutical industry to assess and visualize the effectiveness of cancer prevention medicines. By focusing on pharmaceutical companies like Genentech Inc. (USA), AstraZeneca Pharmaceutical PLC (UK), Boehringer Ingelheim (Germany), and Chugai Pharmaceutical Co. Ltd. (Japan), this study seeks to evaluate which cancer-preventing drugs from these companies provide the best efficacy while minimizing side effects for patients. The goal is to assist healthcare professionals (doctors and pharmacists) in making informed decisions about the most suitable medications for cancer prevention, ensuring patient safety and optimal treatment outcomes.*

Design/Methodology/Approach: *This research utilizes Orange software's machine learning and data visualization tools, specifically scatterplot graphs, to analyze complex datasets related to cancer prevention drugs. By using scatterplots to concurrently examine multiple parameters, such as Company Name, Drug Class, Medicine Name, Prevention of Cancer Diseases, and Side Effects Percentage the study aims to identify patterns and correlations that can help pharmaceutical companies and healthcare professionals assess drug efficacy and safety. The approach involves analyzing the relationship between drug characteristics and side effects, providing actionable insights into how different treatments might interact with patient health conditions.*

Findings/Results: *The findings suggest that Orange's scatterplot visualizations provide valuable insights into the effectiveness of various cancer prevention medicines across different pharmaceutical companies. By enabling the simultaneous analysis of multiple parameters, the software helps to identify which drugs are most effective in preventing cancer while minimizing side effects. This provides a clearer understanding of the correlations between drug characteristics, prevention outcomes, and side effects, supporting data-driven decision-making in pharmaceutical development and healthcare practices.*

Originality/Value: *The originality of this study lies in the application of Orange's data mining and machine learning capabilities to visualize complex relationships within pharmaceutical datasets. The use of scatterplots to analyze drug efficacy, prevention outcomes, and side effects is an innovative approach that offers a richer, more nuanced understanding of cancer prevention drugs. This study contributes valuable insights into optimizing drug choice and treatment strategies, ultimately improving patient safety and therapeutic outcomes.*

Paper Type: *This is an analytical research paper that applies machine learning and data mining techniques to assess the effectiveness and safety of cancer prevention medicines. The research focuses on using Orange software's visualization tools to extract and interpret complex data, providing actionable insights for pharmaceutical companies and healthcare professionals.*

Keywords: Pharmaceuticals Industries Companies, Lung Cancer Diseases, Drug Classes, Medicines, Orange Software, Scatter Plot and Visualization

1. INTRODUCTION:

Data mining visualization techniques are a set of methods used to represent data graphically, making it easier to comprehend and interpret complex datasets. These techniques benefits in different parties involved, such as business analysts, data scientists, decision-makers, and non-technical users. They help in identifying patterns, correlations, and trends that might not be apparent from raw data. An open-source analyzer for analyzing and visualization data called "Orange" has several features, including scatter plots. Particularly beneficial are activities involving interactive data visualization, automated learning and data extraction, Scatter plot graphical representation can be highly beneficial to pharmaceutical companies in several ways related to medicines for prevention and side effects correlation find out. Genentech Inc USA, AstraZeneca Pharmaceutical PLC.UK, Boehringer Ingelheim Pharmaceutical Inc. Germany, Chugai pharmaceutical co ltd .Japan and this each of these pharmaceutical companies has contributed to the treatment of non-small cell lung cancer (NSCLC) diseases, Accordingly this all companies of medicines has different prevention diseases and side effects of various, these research paper exploring drugs represent significant advancements in the treatment non-small cell lung cancer targeting specific genetic mutations and pathways that are involved in the growth and spread of cancer cells. Each pharmaceutical company a vital part in developing these therapies and recover outcomes for lung cancer patients worldwide. Demographics-based predictions indicate that the annual number of new cases of cancer will reach 35 million by 2050. (Bray et.al., 2024) [1]. India witness approximately 1,200,000 new cancer cases each year. According to the latest data from the National Cancer Registry, 1 in 8 men and 1 in 9 women in India are projected to develop some form of cancer during their lifetime. The lung cancer diseases prevented with new treatment of medical technologies. Nowadays several drugs available in the medical markets besides diagnostic methods have emerged to control respective cancer and have assisted in curing this disease to some extent. The pharmacy companies' production of medicines for lung cancer diseases prevention and side effects percentages analysis with scatter plot graphical representation methods explore the different database information at same time, user can be view the all database information at the same time. This scatter plot visualization method helpful to the Pharmaceutical companies of improvement of efficiency in drugs prevention and reduce side effects diseases.

2. OBJECTIVES:

To view and exploring the pharmacy company name, drugs name, drugs class name, prevention cancer diseases and minimum side effects of cancer diseases medicine name information simultaneously

To promote the awareness in pharmaceutical companies and healthcare department about the cancer diseases prevention medicine's and side effects of multiple datasets views and analysis in scatter plot data mining visualization technique.

3. REVIEW OF LITERATURE RELATED WORKS:

Pharmaceutical companies use various software solutions to store and manage information related to medicines, including details on drug formulations, manufacturing processes, inventory, and distribution.

In the pharmaceutical industry, companies use specialized database software to store, manage, and evaluate a large variety of details. This data includes drug research and development, clinical trials, manufacturing processes, quality control, inventory management, and regulatory compliance. These are a few frequently utilized database software solutions for pharmaceutical industry. (Sarkar.,2023) [2] and ((Kiriiri, et al., 2020) [3]. Lung cancers: Non-small cell lung cancer and small cell lung cancer are the two primary forms. This is identified based on how the cells look under a microscope: NSCLC diseases common type of lung cancer, comprising of approximately 80% of lung cancers. It is less aggressive than SCLC. If discovered early, surgery and/or radiation therapy, chemotherapy may offer a chance of cure. **Small cell lung cancer (SCLC)** SCLC is fast-growing and rapidly spreads through by bloodstream and lymphatics to other parts of the body. It is often advanced at diagnosis. It is usually treated with chemotherapy only or in grouping with radiotherapy. (Nooreldeen, & Bach, 2021) [4] and (Chen et al.,2024) [5]. Drug classes: Drug classes are designed to target specific mechanisms in lung cancer cells to inhibit their growth and survival analysis. (Sun et al.,.2021) [6]. Tyrosine kinase inhibitor, Epidermal Growth Factor Receptor (EGFR) inhibitor, Kinase inhibitor are primarily used for treat non-small cell lung cancer NSCLC (Liu, et al., 2017) [7] and (Chen et al., 2022) [8]. The medications are indeed prescribed for treatment of lung cancer: Erlotinib (Tarceva): This drug is an EGFR (epidermal growth factor receptor) inhibitor. It's used primarily for non-small cell lung cancer (NSCLC) that is locally advanced or metastatic and has specific mutations in the EGFR gene. et.al (He et al., 2021) [9]. Rozlytrek (Entrectinib): Entrectinib is tyrosine kinase inhibitor that targets ROS1 and NTRK gene fusions. It’s used for NSCLC that has these specific genetic alterations.Gefitinib (Iressa): Similar to erlotinib, gefitinib is an EGFR inhibitor. It's used for NSCLC with certain EGFR mutations.Osimertinib (Tagrisso): This is a third-generation EGFR inhibitor. It's used for NSCLC with EGFR mutations, particularly aimed at those who have developed resistance to first-generation EGFR inhibitors.et.al (Araghi et al., 2023) [10]. Afatinib (Gilotrif): Afatinib is another EGFR inhibitor, but it also targets other HER family receptors. It’s used for NSCLC with specific EGFR mutations. (Han et al., 2021) [11]. Alecensa (Alectinib): This drug is an ALK (anaplastic lymphoma kinase) inhibitor. It’s used for NSCLC with ALK gene rearrangements, often in cases where crizotinib (another ALK inhibitor) is not effective. These targeted therapies are tailored to specific genetic mutations or alterations found in the cancer cells, offering more personalized and possibly real treatment options for patients with lung cancer. (Niu et al.,,.2023) [12] and (Buszka et al., .2022) [13]. MS sql Server widely used for managing clinical trial data, manufacturing records, and inventory. Provides tools for querying, reporting, and data integration. (Savoska, & Ristevski,,.2020) [14]. **Oracle Database:** Offers robust data management and security features, commonly used for storing research data, regulatory documents, and manufacturing information. (Bosilj-Vukšić & Spremić, 2005) [15] and (Subramanian et al.., 2021) [16]. **MySQL:** An open-source RDBMS utilized in a number of applications, such as managing laboratory data and smaller-scale projects. **MongoDB:** Used for storing and managing unstructured data, such as research notes and experimental data. Supports flexible schema design. (Matallah et al., 2021) [17]. **Cassandra:** A distributed NoSQL database utilized to manage substantial amounts of data with high availability, suitable for real-time data applications like pharmacovigilance. **SAP ERP:** Manages data related to production, inventory, and supply chain operations. Includes modules for financials and compliance reporting’s (Ali et al.2023) [18]. **Data views and analysis:** Data mining software comes with various data views and analysis features that help users uncover insights from large datasets. Such that data base of data views and analysis types. **a. Data Exploration:** This is the initial step where users explore the data to understand its structure, quality, and basic statistics. **Tableau and Microsoft Power BI:** The software’s are used to statistical applications, data distribution, visualizations of histograms and scatter plots. (Molke et al., 2024) [19]. **b. Descriptive Analytics:** This involves summarizing historical data to understand what has happened. **SAS and IBM SPSS:** This software are used to aggregated reports, dashboards showing trends over time and descriptive statistics. (Huang et al., .2024) [20]. **c. Predictive Analytics:** This type of analysis uses historical data to make predictions about future events.

RapidMiner and KNIME: This software are used to Predictive models such as regression analysis, decision trees, and machine learning algorithms .(Ogunleye et al., 2019)[21].**d**. **Diagnostic Analytics:** This focuses on understanding why something happened by analyzing past data. **QlikView and Looker:** This software are used to Drill-down features to investigate anomalies or patterns, correlation analysis to find relationships between variables, and root cause analysis .(Arief, et al., 2023)[22]. e**. Prescriptive Analytics:** This recommends actions based on the data analysis to optimize outcomes. IBM Watson Analytics and Microsoft Azure Machine Learning. This software are used to optimization models, scenario analysis and recommendations for resource allocation. (Kumar et al.,2022) [23] and (Harfoushi & Hasan, 2018) [24] .**Text Mining:** It entails textual data analysis to extract useful information and insights. KNIME and RapidMiner. This is software used to Sentiment analysis and topic modeling (Sakai et al.,.2024).[25].Association Rule Learning. This technique can be applied to identify correlations between variables in large datasets. **Orange and Weka**: This software are used to Market basket analysis to identify items frequently bought together. (Ratra & Gulia, ,2020) [26]. **Cluster Analysis:** In order to accomplish data points must be grouped into clusters based on similarities.R (with packages like k-means) and Python (with scikit-learn). This software are used to Segmentation of consumers into separate groups based on purchasing behavior, or clustering of geographical regions with similar sales patterns. (Lüdecke et al., 2021) [27] and (Lone & Warale,, 2022)[28]. .**Anomaly Detection:** This identifies unusual data points that do not conform to the expected pattern. **SAS and MATLAB.** This software are used to Detection of fraudulent transactions or system anomalies by analyzing deviations from the norm in transactional data or system logs .(Sonkar, et al ., 2019)[29].**Visualization:** This involves creating graphical representations of data to uncover insights and patterns. **Tableau and Power BI:** This software's used to Interactive dashboards, heat maps, time series charts, and geographical maps to visualize data trends and patterns.(Raihen, M. N. et al, 2025) [30]. **Orange software:** Scatter plot visualization techniques useful to simultaneously view and analyze multiple datasets in a unified scatter plot, making it easier to compare and interpret the data. Orange facilitates comprehensive exploration of multidimensional data through its graphical representations and integrated analysis tools. Its ability to handle complex datasets and provide intuitive visualizations makes it a powerful tool for researchers, analysts, and data scientists aiming to uncover insights from multidimensional data sources. l(Dobesova, Z. 2024)[31].Data mining and visualization techniques are increasingly used in cancer research to identify patterns, predict outcomes, and support early detection. Techniques like clustering, decision trees, and neural networks help analyze large genomic and clinical datasets, uncovering insights about cancer subtypes, progression, and treatment response. Visualization tools such as heatmaps, survival curves, and 3D molecular structures allow researchers to better interpret complex data and present findings in a more accessible manner.

4. PROBLEM OF THE STATEMENTS:

In the pharmaceutical industry, efficiently identifying and viewing multiple data sets simultaneously can be challenging, because Pharmaceutical companies of production of data information is a Complex ity of Data Relationships. Pharmaceutical data can be complex, with numerous interrelated variables, applications that does not support advanced data querying or visualization may struggle to present this information effectively.

5. NEED OF THIS STUDY:

Predictive Modeling: Advanced analytics in pharmaceuticals often involves predictive modeling advanced tools and approaches, pharmaceutical companies can better manage the complexity of their documents, enabling more effective analysis and decision-making.

6. RESEARCH METHODOLOGY:

Research methodology is the systematic approach to solving a research problem. It involves identifying, selecting, processing, and analyzing data to provide accurate and reliable results.

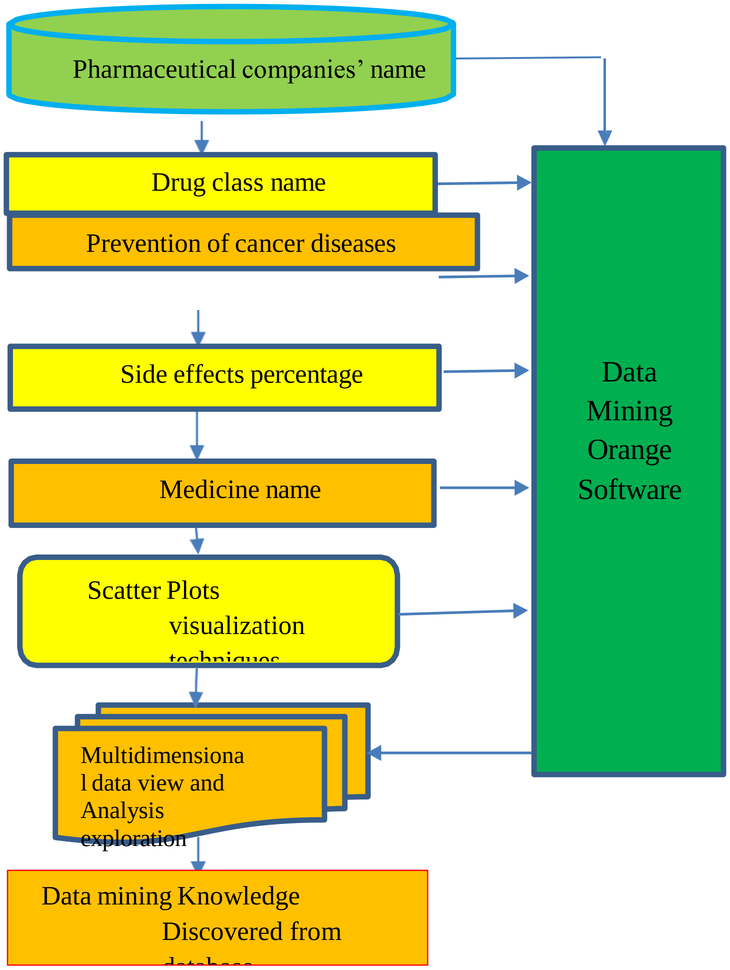


Figure: 1 Data Mining Knowledge Discovery Techniques Processing for Views the Multiple Datasets

7. STUDY DESIGN:

Data mining knowledge discovery techniques in pharmaceutical companies improve drug discovery, improve clinical trials, improve safety monitoring, inform market strategies, and support personalized medicine. This research paper explores the pharmaceutical industries production of drug class, medicine, prevention of cancer diseases, side effects percentage of information all details are view and analysis simultaneously using with Orange software scatter plot methods.

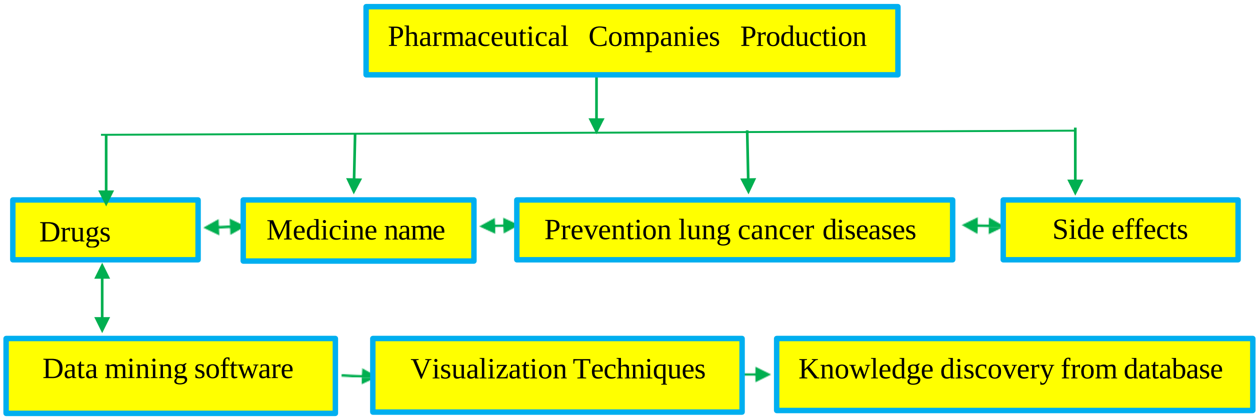


Figure 2: Study Design Process in Pharmaceutical Industries Production of Drug Class's Aspects and Multiple Datasets Views and Analysis

7.1 Pharmaceutical Companies:

There are many pharmaceutical companies involved in lung cancer treatment, each with its own portfolio of drugs and research. keeping track of the latest developments and drugs from multiple companies can be overwhelming. (Hot et al., 2019)) [32] and (Ogura et al., 2023) [33].

7.2 Drug Classes:

Drugs are classified based on their mechanisms of action (e.g., chemotherapy, targeted therapy, immunotherapy). Each class has multiple drugs with different indications and side effects (Gelatti et al., 2019) [34] and (Jassim, M. M., & Jaber, M. M. et al, 2022) [35].

7.3 Medicines Details:

Each medicine has specific indications, dosage forms, administration routes, side effects, interactions, and patient considerations. Managing detailed information about each drug can be complex. (Honeywell et al., 2020) [36].

7.4 Lung Cancer Diseases:

Lung cancer has various subtypes (NSCLC, SCLC) and each subtype can have different characteristics and treatment options. Understanding the nuances of these subtypes requires extensive knowledge. NSCLC common and generally grows more slowly, while SCLC tends to grow and spread more quickly. (Nooreldeen & Bach, 2021) [37].

7.5 Lung Cancer:

Lung cancer is the name for cancers that start in lungs Lung cancer usually starts in the airways (bronchi or bronchioles) or small air sacs (alveoli) of your lungs. It can then spread to other organs (Onoi et al., 2020) [38]. "Lung cancer" occurs when abnormal cells in the lung tissue grow uncontrolled ably and form tumors. These cancerous cells can invade nearby tissues and may spread (metastasize) to other parts of the body .Such as the brain, liver, or bones. (www.my.clevelandclinic.org)

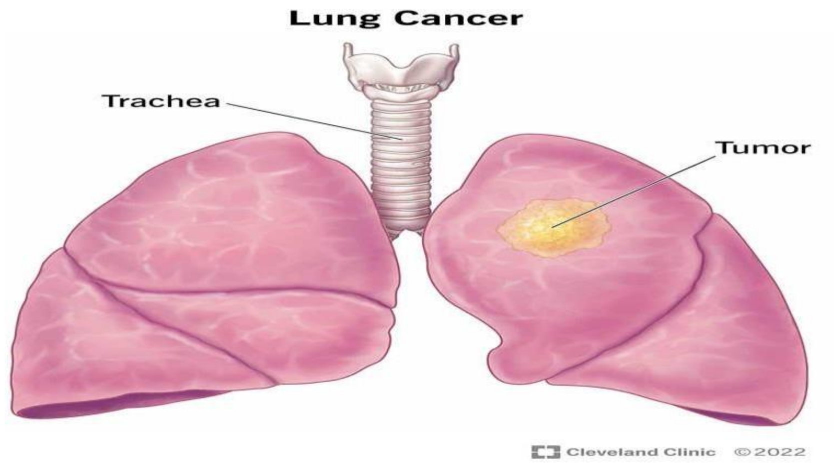


Figure 3: Lung Cancer Diseases

7.6 Small Cell Lung Cancer:

A Small Cell Lung Cancer (SCLC) is a type of lung cancer characterized by small, round cells that grow quickly and form large tumors. SCLC is a rare fast-growing lung cancer. Small cell lung cancer can affect anyone in this world (Zugazagoitia, et al.,.2022)[39].and (www.my.clevelandclinic.org)

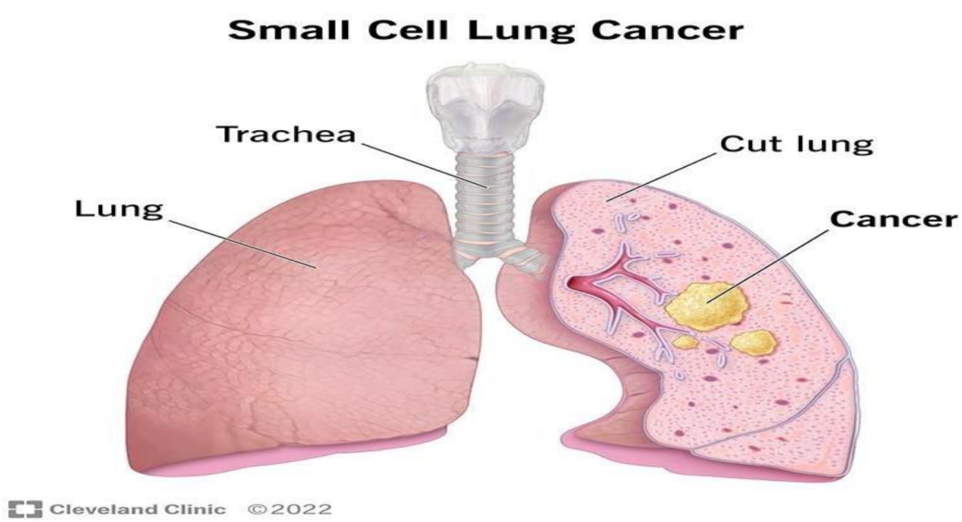


Figure 4: Small Cell Lung Cancer Diseases

7.7 Non - Small Cell Lung Cancer:

Non-Small-Cell Lung Cancer (NSCLC) is the most common type of lung cancer, making up about 85% of cases. Knowing the stage of the cancer is crucial as it helps determine the most effective treatment approach. The tumor's size determines the NSCLC stages., whether cancerous cells have spread to lymph nodes, and whether it has metastasized to other parts of the body. (Daly et al., 2022)[40] and (www.zenonco.io)

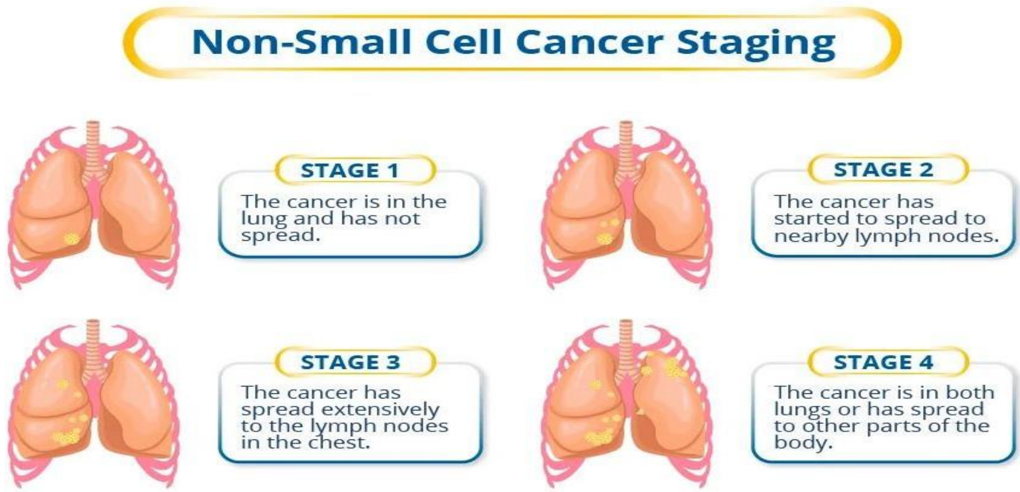


Figure 5: Small Cell Lung Cancer Diseases

Stage 0: Also known as **carcinoma in situ**, stage 0 NSCLC is very early cancer that is only present in the outermost layers of cells lining the lungs. At this stage, the cancer hasn't spread deeper into lung tissues or outside the lungs.

Stage I:NSCLC is localized, meaning it has not spread to the lymph nodes. It is divided into two sub stages based on tumor size:

Stage IA: The tumor is 3 centimeters (cm) or smaller.

Stage IB: The tumor is larger than 3 cm but not larger than 4 cm.

Stage II:NSCLC is characterized by larger tumors or cancer that has begun to spread to nearby lymph nodes. It is subdivided into:

Stage IIA: The tumor is larger than 4 cm but not larger than 5 cm, or cancer has spread to nearby lymph nodes on the same side of the chest.

Stage IIB: The tumor is larger than 5 cm but not larger than 7 cm, or there are smaller tumors in the same lobe of the lung.

Stage III:NSCLC is more advanced cancer that has spread to the lymph nodes and possibly to nearby structures and organs. It is divided into three sub-stages:

Stage IIIA: Cancer has spread to the lymph nodes on the same side of the chest but not to distant parts of the body.

Stage IIIB: The tumor may be any size and cancer has spread to lymph nodes near the collarbone or opposite side of the chest.

Stage IIIC: Cancer has spread to the chest wall, diaphragm, or other nearby structures, making it more challenging to treat.

Stage IV: The most advanced stage of NSCLC, Stage IV, indicates that cancer has metastasized, or spread, to both lungs, to the fluid around the lungs or heart, or to distant parts of the body such as the liver, bones, or brain. This stage is further divided into:

Stage IVA: Cancer has spread within the chest and/or has metastasized to one distant site.

Stage IVB: Indicates a more extensive spread to multiple parts of the body.

Understanding the stage of Non-Small-Cell Lung Cancer is vital for determining the most appropriate treatment plan. Treatments may include surgery, chemotherapeutic, radiation therapy, targeted therapy, or a combination of these approaches, depending on the stage and other factors related to the patients' health. Always consult with a medical professional for the most accurate diagnosis and treatment options. (Petrella, et al., 2023) [41].

7.8 Pancreatic Cancer:

Pancreatic cancer is a type of malignancies that begins in the tissues of the pancreas, an organ located behind the lower part of the stomach. Blood sugar regulation and digestion is greatly aided by the pancreas. The prognosis for pancreatic cancer is often poor due to late-stage diagnosis. The survival rates vary based on the stage at diagnosis, overall health, and response to treatment. Early detection and advances in treatments are crucial for improving outcomes and pancreatic cancer is classified into stages based on the extent of the disease, which helps guide treatment decisions and predict outcomes.

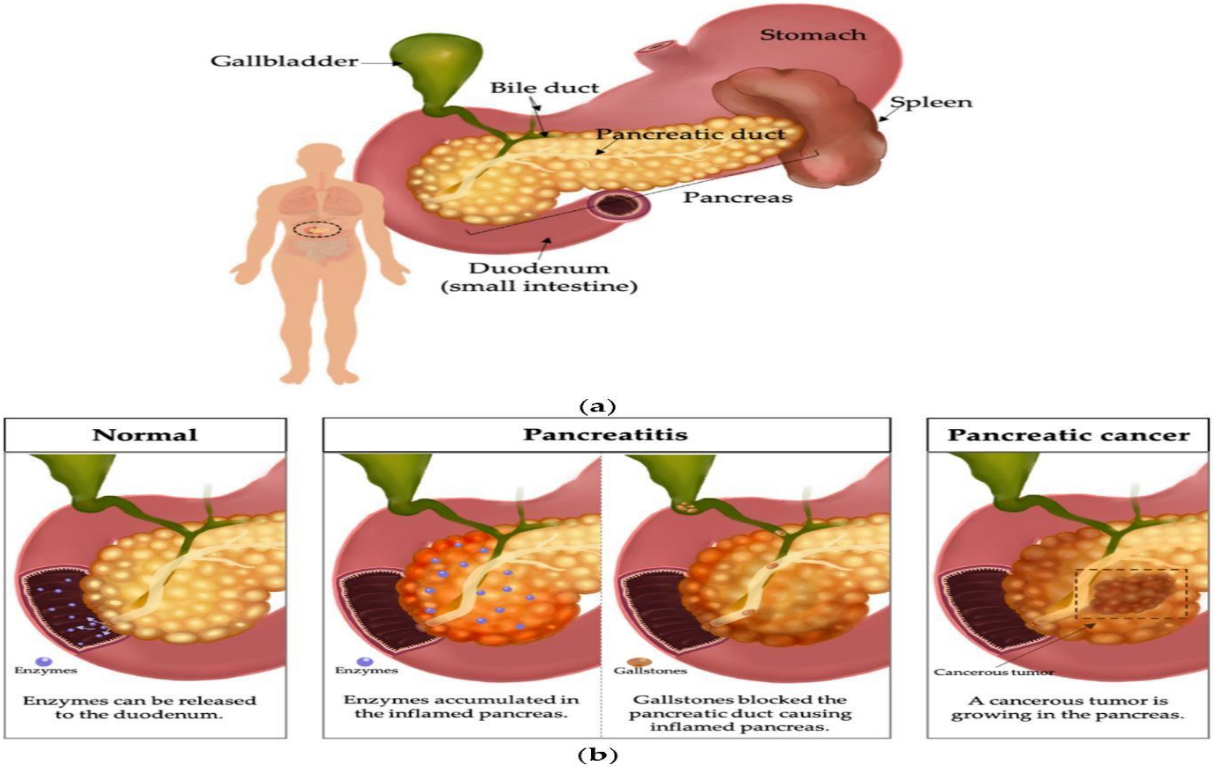


Figure 6: Pancreatic Cancer Diseases

Illustration of the pancreas. (a) Location of the pancreas in the human body, (b) comparison among the normal pancreas, pancreatitis, and pancreatic cancer (Ungkulpasvich, et al.., 2023) [42] and (Cao et al.,2021) [43]. The prognosis for pancreatic cancer is often poor due to late-stage diagnosis. The survival rates vary based on the stage at diagnosis, overall health, and response to treatment. Early detection and advances in treatments are crucial for improving outcomes and pancreatic cancer is classified into stages based on the extent of the disease, which helps guide treatment decisions and predict outcomes.

7.9 Side Effects of Percentages:

The drugs classes and medication has a low or high incidence of severe side effects; it may not be the best choice for certain patients. Knowing these percentages helps in selecting medications with a better safety profile for the patient's specific needs. Healthcare professional are understanding medicine of side effects percentage, including how common they are, helps in managing potential drug interactions. For instance, combining drugs with similar side effects may exacerbate issues, so healthcare professionals can adjust treatment plans accordingly. Eroletinib (Tarceva), Rozlytrek (Entrectinib), Gefitinib (Iressa), Osimertinib (Tagrisso), Galotti (afatinib), Alecensa (alectinib) all the medicines of side effects percentage retrieved from website (www.goodrx.com).

7.10 Data Mining Software:

Orange has tremendous open source data mining and visualization tool that can be highly beneficial for pharmaceutical companies, doctors, and pharmacists. Such as (a) **Drug Discovery and Development:** Pharmaceutical companies can use Orange to analyze large datasets from clinical trials, identify potential drug targets, and optimize drug formulations.(b).**Patient Data Analysis:** Doctors and pharmacists can use Orange to analyze patient data, identify trends, and make data-driven decisions for personalized treatment plans.(c).**Market Analysis**: Pharmaceutical companies can use Orange to analyze market trends, understand the competitive landscape, and develop effective marketing strategies.(d) **Predictive Analytics**: Orange can help in predicting disease outbreaks, patient outcomes, and the effectiveness of treatments, which is crucial for both healthcare providers and pharmaceutical companies. (Siregar, et al., 2023) [44].

7.11 Visualization Techniques:

Orange provides intuitive visualizations that help in understanding complex data, making it easier for healthcare professionals to interpret and act on the data. Orange Data Mining software offers variety of visualizations, including scatter plots, which are particularly useful for displaying relationships between multiple variables and it can also incorporate a variable by adjusting the color, size, or shape of the data points. Scatter plots visualization implemented by methods of **(a) Regression Analysis:** This is used to fit a line through the data points to show trends or relationships between variables. **(b). Principal Component Analysis (PCA):** PCA is used to decrease the dimensionality of the data, making it easier to visualize multiple variables on a 2D scatter plot. (**c). Clustering Algorithms**: These can be used to identify and visualize clusters or groups within the data. (Shrivastava et al., 2023) [45].

7.12 Knowledge Discovery from Database:

KDD is the process of discovering useful knowledge from large volumes of data. It involves several steps, from data collection to interpretation, and often involves using data mining techniques to extract meaningful patterns. KDD helps the Pharmaceutical Company transform raw transactional data into actionable insights. By understanding purchasing patterns and healthcare professional segments, the company can make data-driven decisions to enhance sales, optimize product placement, and improve marketing strategies. (Ranjan, 2009) [46] and (Gillespie et al, 2019) [47].

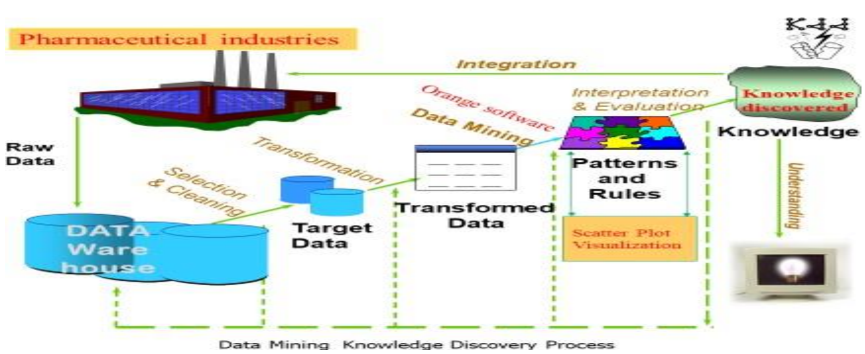


Figure 7: Data Mining Scatter Plot Visualization Techniques Process

Data mining is a process used to discover patterns, correlations, extraction of useful information from large sets of data.

7.13 Unprocessed data:

Raw data refers to the unprocessed data collected from various sources. This data is often unstructured and may include text, numbers, dates, or other formats.

7.14. Data warehouse:

A data warehouse is a centralized repository that stores integrated data from multiple sources. It is designed to support querying and analysis, and it typically involves historical data.

7.15 Selection and Cleaning:

Selection: Selecting the relevant data involves choosing a subset of data that is pertinent to the specific analysis or problem.

7.16 Transformation:

Transformation involves converting data into a format suitable for analysis. This step prepares data for mining by making it more suitable for algorithms.

7.17 Data Mining:

Data mining is the core step where algorithms are applied to the prepared data to discover patterns, relationships, or insights.

7.18 Interpretation and Evaluation:

Interpretation: Interpreting the results involves understanding the patterns or insights discovered through data mining and Evaluation: Evaluating the performance and utility of the data mining models and findings.

7.19 Knowledge Discovered:

Knowledge discovered refers to the actionable insights and patterns identified by data mining that can be used to make informed decisions or drive actions. These steps all organizations can effectively mine data to uncover valuable insights and drive decision-making in efficiently. (Odoh et al., 2024) [48] and (Qi et al., 2024) [49].

8. METHODOLOGY AND ORANGE SOFTWARE:

Orange is a popular open-source data mining and machine learning software suite that offers various tools for data analysis and visualization. One of the powerful features in Orange is its ability to visualize data with scatter plots, which can help in understanding relationships between variables and Orange software can efficiently create scatter plots that visualize relationships between multiple variables simultaneously. This allows for an in-depth understanding of multiple datasets relationships and decision making it easier to identify patterns, trends, and anomalies. (Estupiñán Ricardo. et al., 2021) [50] and (Novianto, & Triraharjo, 2024) [51]

9. ORANGE SCATTER PLOT INTEGRATIVE EXPLORATION

Pharmaceutical industries production of drugs classes, medicines, lung cancer diseases prevention and side effects dataset's views them simultaneously with Orange software scatter plots techniques.

a. The Following Steps:

Step 1: **Data Import:** To import the dataset into Orange first prepare the file format CSV or Excel.
Step 2: **File Exploration:** To open the CSV file database through File widget
Step 3: **Configure the Plot:** To connect the data to the scatter plot widget, drag the scatter plot

widget onto the workflow canvas

Step 4: **Configure Axes:** Set the X-axis to "Variable" and Set the Y-axis to "Variable".

Step 5: **Configure attributes: a. Colors:** Differentiate between categories

i. **Shapes:** Highlight different groups

ii. **Sizes:** Indicate magnitude

iii. **Labels:** Provide specific identification and context

Step 6: Scatter Plot visualizations help in understanding relationships and patterns across multiple variables simultaneously

b. Orange Software Work Flow Diagram:

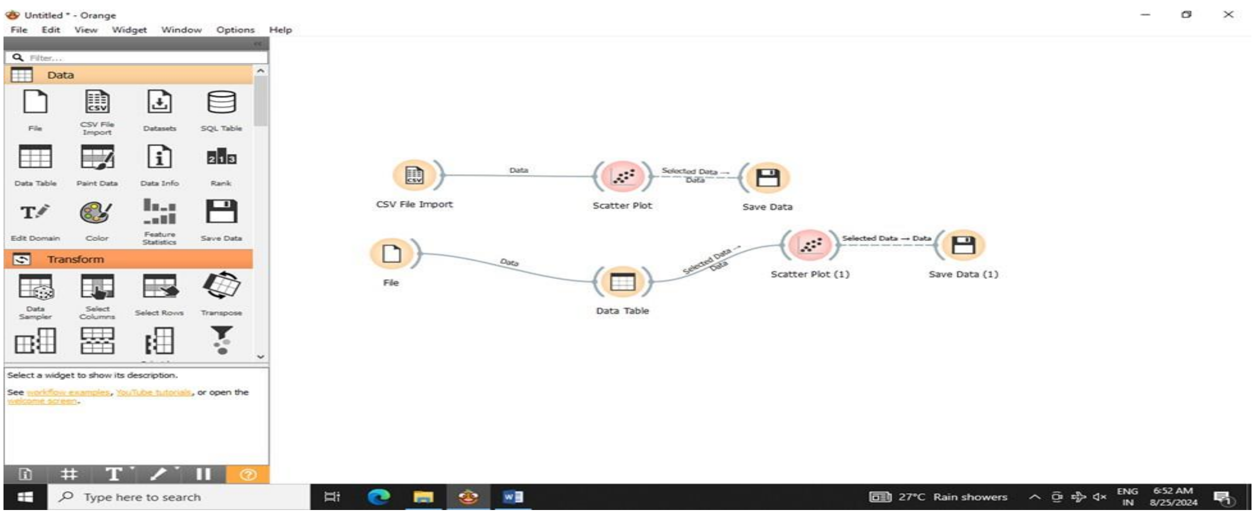
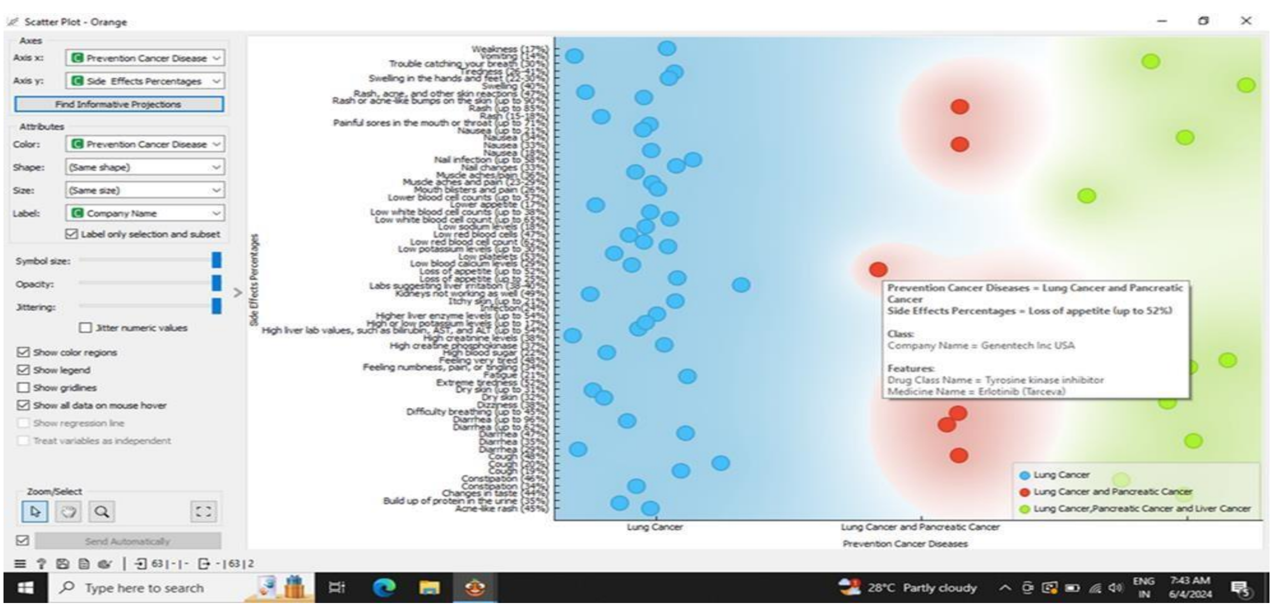


Figure 8: Scatter Plot Representation

c. Scatter Plot:



Graph 1: Prevention of Cancer Diseases vs Side Effects Percentage: Outcome View and Analysis

Orange data mining software of scatter plot is used to visualize the relationship between two variables. Each point on the scatter plot represents an observation from the dataset, with its position determined by the values of two variables: one on the x-axis and one on the y-axis. The basic formula used to plot these points is: (xi, yi), where (x_i) is value of the variable on the x-axis for the (i)-th observation, and (y_i) is the value of the variable on the y-axis for the same observation. x = Prevention of cancer diseases. y = Side Effects Percentage of search information displayed the below view structure.

Prevention Cancer Diseases = Lung cancer and Pancreatic cancer
Side Effects Percentage = Nausea (33%)
Class
Company Name = Genentech Inc USA
Feature
Drug class Name = Tyrosine kinase inhibitor
Medicine Name = Erlotinib(Tarceva)

10 . DATA COLLECTION:

The process of obtaining information from diverse sources in order to meet particular research questions and objectives is referred to as data collection. The doctors and pharmacists has difficulty to simultaneously manage and view comprehensive information about pharmaceutical companies, lung cancer diseases, drug classes, and specific medicines. This difficulty is rooted in several factors related with complexity of the data and the multidimensional nature of the information. This research paper has been explored the knowledge about the complexity of information such as **Pharmaceutical Company Name:** The name of the company that develops, manufactures, and markets medications. **It refers to** Companies conduct research, clinical trials, and regulatory submissions to bring new drugs to market. **Drug Class Name:** A category of drugs that share similar chemical structures, mechanisms of action, or therapeutic effects. **Role:** Helps categorize drugs with similar properties or effects, making it easier to understand their uses and interactions. **Medicine Name:** The specific name of a drug or medication, often a brand name or generic name. **Role:** Identifies individual drugs with unique formulations and indications. **Prevention Diseases:** Refers to diseases that a drug or medication prescribed to prevent or manage. **Role:** Defines the therapeutic scope of the medication and its intended use in preventing disease progression or occurrence. **Side Effects Percentages:** The frequency and likelihood of adverse effects associated with a medication, usually expressed as a percentage. **Role:** Provides information on the safety profile of a drug, helping healthcare providers weigh benefits against potential risks. Accordingly, These database structure view with Data mining Orange software scatter plot visualization techniques views across multiple variables simultaneously. The pharmaceutical companies (1). Genentech Inc USA, (2). AstraZeneca Pharamceuticals PLC UK, (3). Boehringer Ingelheim Pharmaceuticals Inc Germany, (4). Chugai Pharmaceutical Co Ltd Japan, these all the companies productions of details such as Company Name, Drug Class Name, Medicine Name, Prevention Cancer Diseases and Side Effects Percentages views on the Orange software scatter plot visualization technique emphasizes interactive data exploration, Users can be dynamically select subsets of data from X axis and Y axis datasets view with zoom in a specific areas of interested in plots and interactively adjust parameters to visualize how different variables affect each other.

This interactively is crucial for exploring complex relationship in multidimensional datasets. The Table:1 and Table:2 Genentech Inc USA, Table:3 and Table:4 AstraZeneca Pharmaceuticals PLC UK, Table:5 Boehringer Ingelheim Pharmaceuticals Inc Germany, Table: 6 Chugai Pharmaceutical Co Ltd Japan details are represented below table structures.

Table 1: Genentech Inc USA.

Company Name	Drug Class Name	Medicine Name	Prevention Cancer Diseases	Side Effects Percentages
Genentech Inc USA	Tyrosine kinase inhibitor	Eroletinib (Tarceva)	Lung Cancer and Pancreatic Cancer	Rash (up to 85%)
Target: Epidermal Growth Factor Receptor (EGFR) **Indications:** NSCLC with specific EGFR mutations, pancreatic cancer **Mechanism:** Inhibits the tyrosine kinase domain of EGFR, blocking downstream signaling pathways that drive tumor growth.				Diarrhea (up to 62%)
				Loss of appetite (up to 52%)
				Extreme tiredness (52%)
				Cough (48%)
				Difficulty breathing (up to 45%)
				Nausea (33%)
				Infection (24%)

Table 2: Genentech Inc USA

Company Name	Drug Class Name	Medicine Name	Prevention Cancer Diseases	Side Effects Percentages
Genentech Inc USA	Tyrosine kinase inhibitor	Rozlytrek (Entrectinib)	Lung Cancer,Pancreatic Cancer and Liver Cancer	Lower blood cell counts (up to 57%)
Target: ROS1 and NTRK fusions **Indications:** NSCLC with ROS1 rearrangements, other cancers with NTRK fusions **Mechanism**: Targets ROS1 and NTRK gene fusions, inhibiting their tyrosine kinase activity and blocking tumor growth.				Feeling very tired (48%)
				Constipation (46%)
				Changes in taste (44%)
				Swelling (40%)
				Dizziness (38%)
				Nausea (34%)
				Feeling numbness, pain, or tingling (34%)
				Trouble catching your breath (30%)
				Trouble catching your breath (30%)

The above Table 1 and Table 2 represented about the Genentech Inc, USA Company offers two tyrosine kinase inhibitors, Erlotinib (Tarceva) and Entrectinib (Rozlytrek), which are used to treat lung and pancreatic cancers and these medications represent significant advancements in targeted cancer therapy, offering effective treatment options with relatively manageable side effects.

Table 3: AstraZeneca Pharmaceuticals PLC UK

Company Name	Drug Class Name	Medicine Name	Prevention Cancer Diseases	Side Effects Percentages
AstraZeneca Pharamceuticals PLC UK	Epidermal Growth Factor Receptor (EGFR)	Gefitinib (Iressa)	Lung Cancer	Rash, acne, and other skin reactions (47%)

	inhibitor			
Target: EGFR **Indications:** NSCLC with EGFR mutations **Mechanism:** Selectively inhibits the tyrosine kinase domain of EGFR, blocking signaling pathways involved in cell proliferation.				Labs suggesting liver irritation (38-40%)
				Buildup of protein in the urine (35%)
				Diarrhea (29%)
				Nausea (18%)
				Vomiting (14%)
				Weakness (17%)
				Lower appetite (17%)

Table 4: AstraZeneca Pharamceuticals PLC UK

Company Name	Drug Class Name	Medicine Name	Prevention Cancer Diseases	Side Effects Percentages
AstraZeneca Pharamceuticals PLC UK	Epidermal Growth Factor Receptor (EGFR) inhibitor	Osimertinib (Tagrisso)	Lung Cancer	Low white blood cell count (up to 65%)
Target: EGFR, particularly the T790M mutation **Indications:** NSCLC with EGFR T790M mutation **Mechanism**: Irreversibly binds to the EGFR tyrosine kinase domain, specifically targeting the T790M resistance mutation, leading to inhibition of tumor growth.				Low platelets (53%)
				Low red blood cells (47%)
				Diarrhea (47%)
				Acne-like rash (45%)
				Muscle aches/pain (36%)
				Nail changes (33%)
				Dry skin (32%)
				Mouth blisters and pain (26%)
				Fatigue (21%)
				Cough (20%)

The above Table 3 and Table 4 represented about the AstraZeneca Pharamceuticals PLC UK. Company offers two notable EGFR inhibitors: Osimertinib (Tagrisso) and Gefitinib (Iressa).and these both medicines of treatment of Lung Cancer and minimum side effects. Thereby helping to control the growth and spread of cancer cells.

Table 5: Boehringer Ingelheim Pharmaceuticals Inc Germany

Company Name	Drug Class Name	Medicine Name	Prevention Cancer Diseases	Side Effects Percentages
Boehringer Ingelheim Pharmaceuticals Inc Germany	Kinase inhibitor	Galotti (afatinib)	Lung Cancer	Diarrhea (up to 96%)
Target: EGFR, HER2, and other ErbB family receptors **Indications**: NSCLC with EGFR mutations **Mechanism:** Irreversibly inhibits EGFR and other members of the ErbB family, blocking signaling pathways that contribute to				Diarrhea (up to 96%)
				Rash or acne-like bumps on the skin (up to 90%)

tumor proliferation and survival.	Painful sores in the mouth or throat (up to 71%)
	Nail infection (up to 58%)
	Higher liver enzyme levels (up to 54%)
	Kidneys not working as well (49%)
	Low white blood cell counts (up to 38%)
	Dry skin (up to 31%)
	Low potassium levels (up to 30%)
	Loss of appetite (up to 25%)
	Nausea (up to 21%)
	Itchy skin (up to 21%)

Table 6: Chugai Pharmaceutical Co Ltd Japan

Company Name	Drug Class Name	Medicine Name	Prevention Cancer Diseases	Side Effects Percentages
Chugai Pharmaceutical Co Ltd Japan	Kinase inhibitor	Alecensa (alectinib)	Lung Cancer	Low red blood cell count (62%)
Target: Anaplastic Lymphoma Kinase (ALK) **Indications:** NSCLC with ALK rearrangements **Mechanism:** Specifically inhibits ALK tyrosine kinase, disrupting signaling pathways that promote cancer cell growth.				High liver lab values, such as bilirubin, AST, and ALT (up to 54%)
				Tiredness (26-41%)
				High creatinine levels (38%)
				High creatine phosphokinase (37%)
				Constipation (34%)
				Swelling in the hands and feet (22-30%)
				Low blood calcium levels (29%)
				Muscle aches and pain (23-29%)
				High blood sugar (22%)
				Cough (19%)
				Low sodium levels (18%)
				High or low potassium levels (up to 17%)
				Rash (15-18%)

The above Table 5 represented about the Boehringer Ingelheim Pharmaceuticals Inc Germany. Company offer a kinase inhibitor: Galotti (afatinib) this medicine for treatment of lung cancer, its function blocks signaling pathways that promote tumor growth. and minimum side effects. The above Table 6 represented about the Chugai Pharmaceutical Co Ltd Japan. Company offer a kinase inhibitor: Alecensa (alectinib) this medicine for treatment of lung cancer with anaplastic lymphoma kinase (ALK) gene rearrangements and potential side effects. These resources collected from (www.goodrx.com).

11. RESULT AND DISCUSSION:

Data mining Orange software scatter plot graphical representation as a interactive exploration. Its versatility and user-friendly interface make it a preferred choice for researchers seeking to leverage advanced analytics in their investigations across various disciplines.

a. Drug Class Name vs Medicine Name:

Orange software of scatter plot once configured, the scatter plot will display the points based on chosen variables can interact with the plot to explore multiple datasets of relationships view and analysis at same time. coloring and grouping datasets contain a categorical variable like a label, these can use that variable to color the points differently. subsequently this helps differentiate between the datasets visually and efficiently. To view and analysis the relationship between a drug class name and a medicine name using a scatter plot with the x-axis and y-axis.

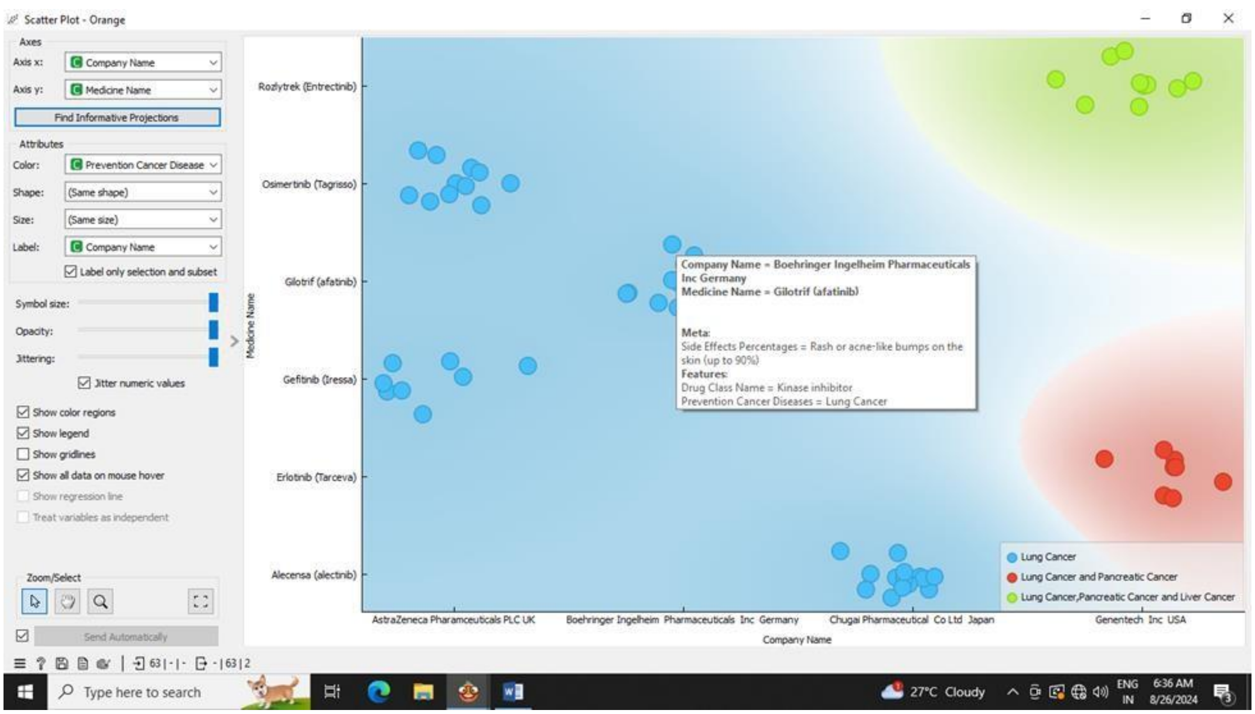
Graph 2: Drug Class Name vs Medicine Name

The graph 2 represented with visually x axis on Drug class name and y axis on Medicine name, these interacted with the scatter plot to explore multiple datasets relationship simultaneously displayed below represented a structured formats.

Drug class Name = Epidermal Growth Factor Receptor (EGFR) inhibitor
Medicine Name= Osimertinib (Tagrisso) Meta
Side Effects Percentage = Cough (20%)
Features
Company Name = AstraZeneca Pharamceuticals PLC UK Prevention
Cancer Diseases= Lung Cancer and Pancreatic Cancer

b. Company Name vs Medicine Name:

To view and analysis the relationship between a company name and a medicine name using a scatter plot with the x-axis and y-axis.



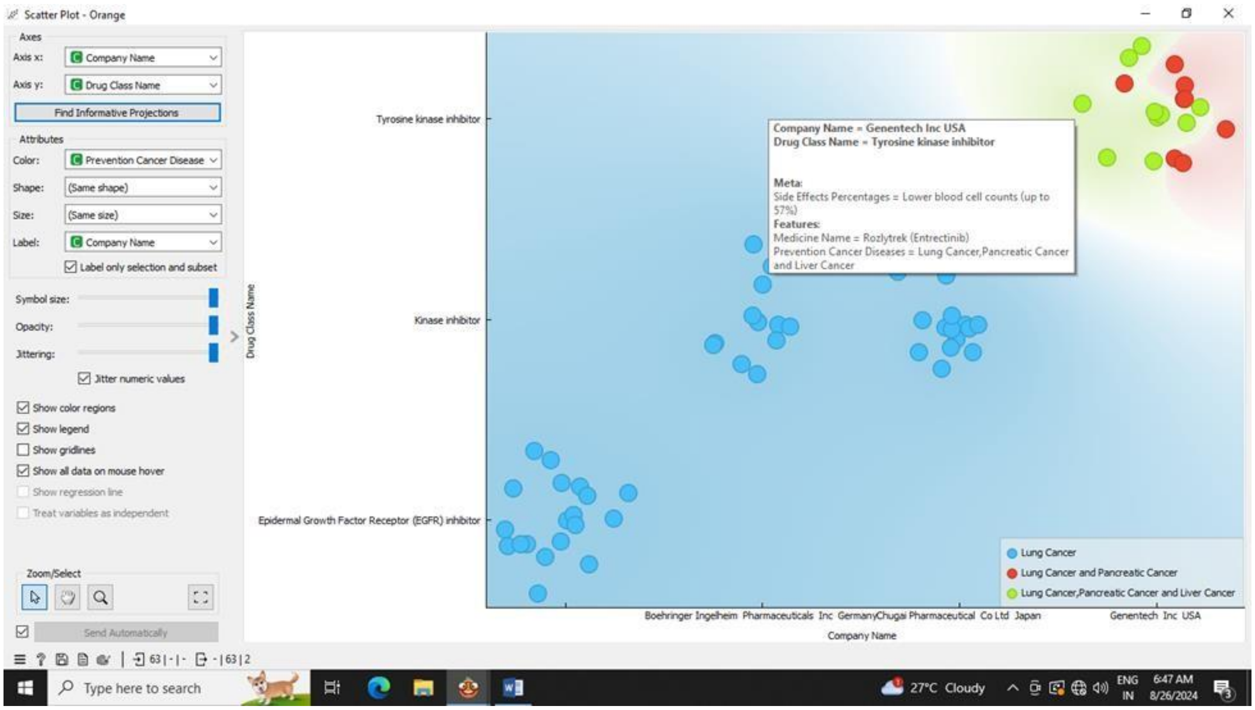
Graph 3: Company Name vs Medicine Name

The graph 3 represented with visually x axis on Company name and y axis on Medicine name, these interacted with the scatter plot to explore multiple datasets relationship simultaneously displayed below represented a structured formats.

Company Name = Boehringer Ingelheim Pharmaceuticals Inc. Germany
Medicine Name = Galotti (afatinib)
Meta
Side Effects Percentage = Rash or acne-like bumps on the skin (up to 90%)
Features
Drug Class Name = Kinase inhibitor
Prevention Cancer Diseases = Lung Cancer

c. Company Name vs Drugs Class Name:

To view and analysis the relationship between a company name and a drug class name using a scatter plot with the x-axis and y-axis.



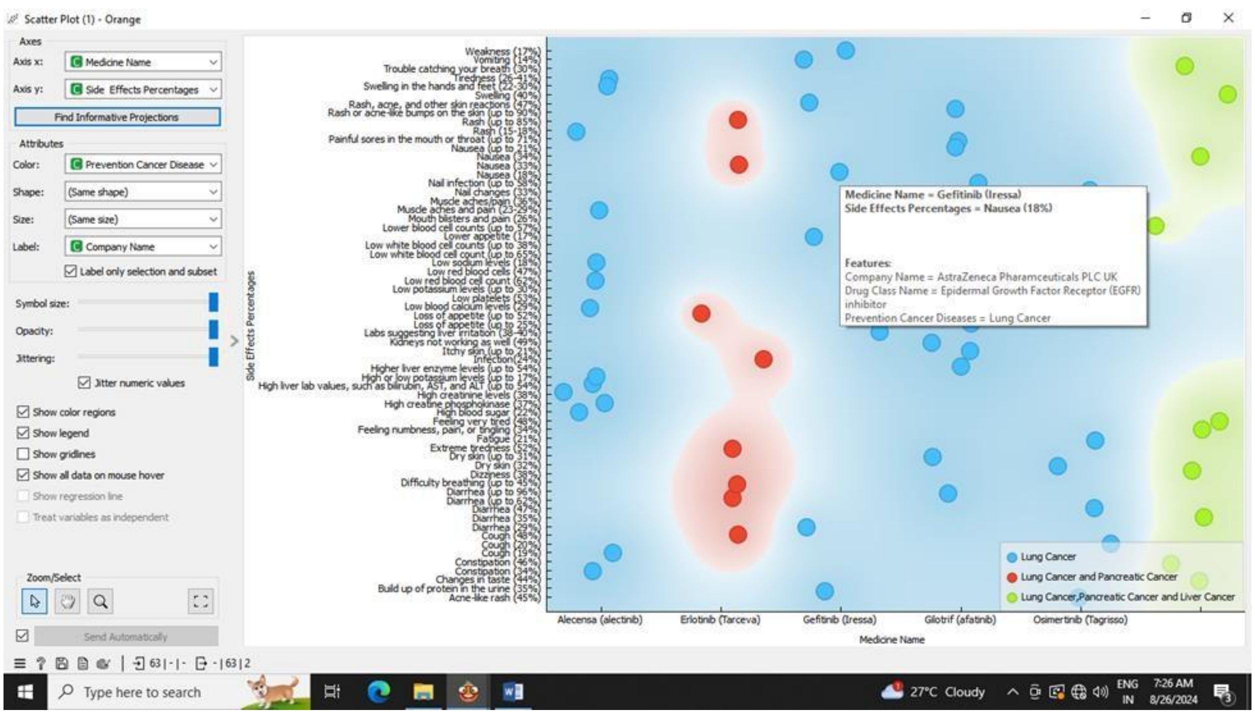
Graph 4: Company Name vs Drugs Class Name

The graph 4 represented with visually x axis on Company name and y axis on Drug class name, these interacted with the scatter plot to explore multiple datasets relationship simultaneously displayed below represented a structured formats.

Company Name = Genentech Inc USA
Drug Class Name =Tyrosine kinase inhibitor
Meta
Side Effects Percentage = Lower blood cell counts (up to 57%)
Features
Medicine Name = Rozlytrek (Entrectinib)
Prevention Cancer Diseases = Lung Cancer ,Pancreatic Cancer and Liver Cancer

d. Medicine Name vs Side Effects Percentages:

To view and analysis the relationship between a medicine name and side effects Percentages medicine name using a scatter plot with the x-axis and y-axis.



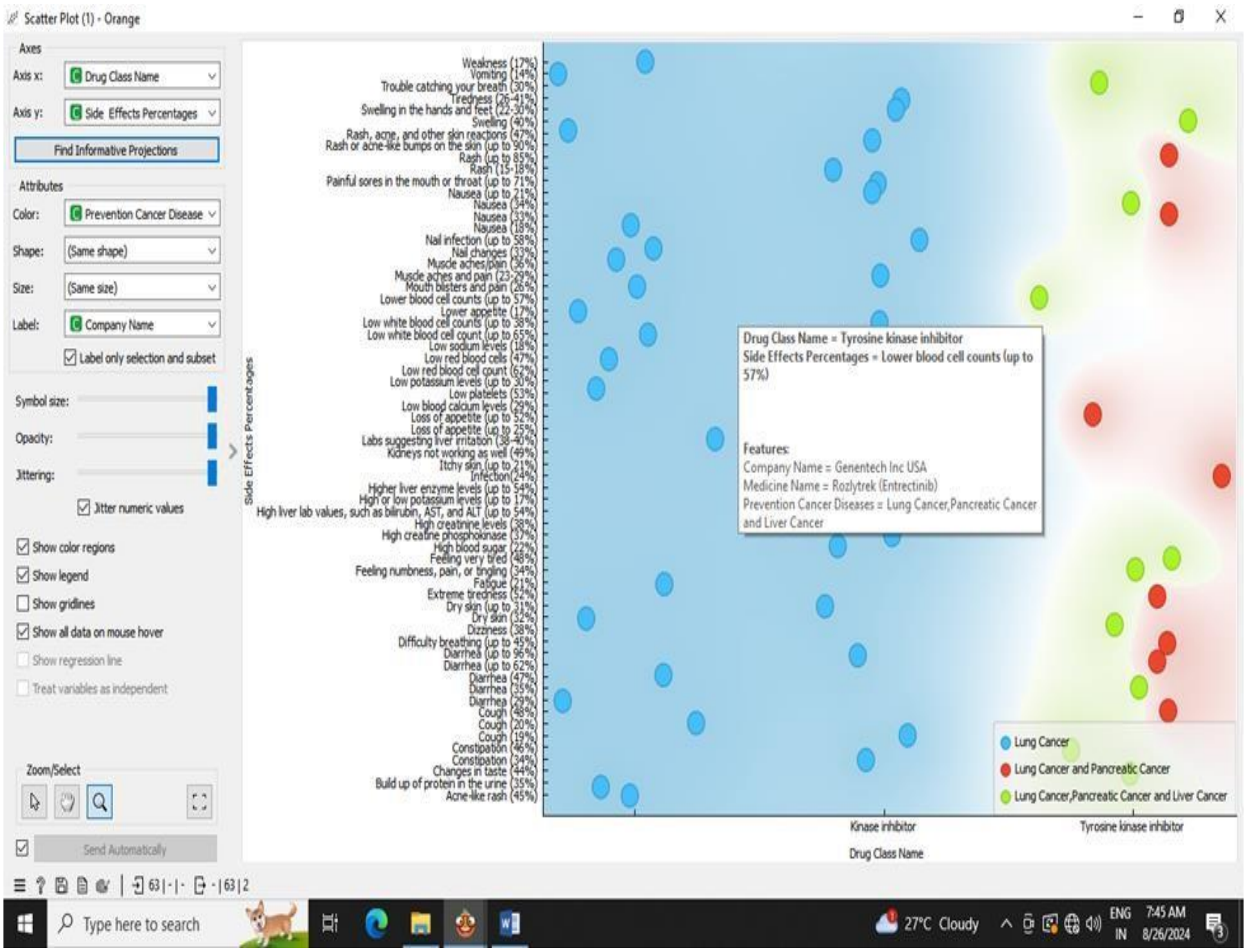
Graph 5: Medicine Name vs Side Effects Percentages

The graph 5 represented with visually x axis on Medicine name and y axis on Side Effects Percentages, these interacted with the scatter plot to explore multiple datasets relationship simultaneously displayed below represented a structured formats.

Medicine Name = Gefitinib (Iressa)
Side Effects Percentage = Nausea
(18%) Features
Company Name = AstraZeneca Pharamceuticals PLC UK
Drug Class Name = Epidermal Growth Factor Receptor (EGFR) inhibitor
Prevention Cancer Diseases = Lung Cancer

e. Drugs Class Name vs Side Effects Percentages:

To view and analysis the relationship between a drug class name and side effects percentages using a scatter plot with the x-axis and y-axis.



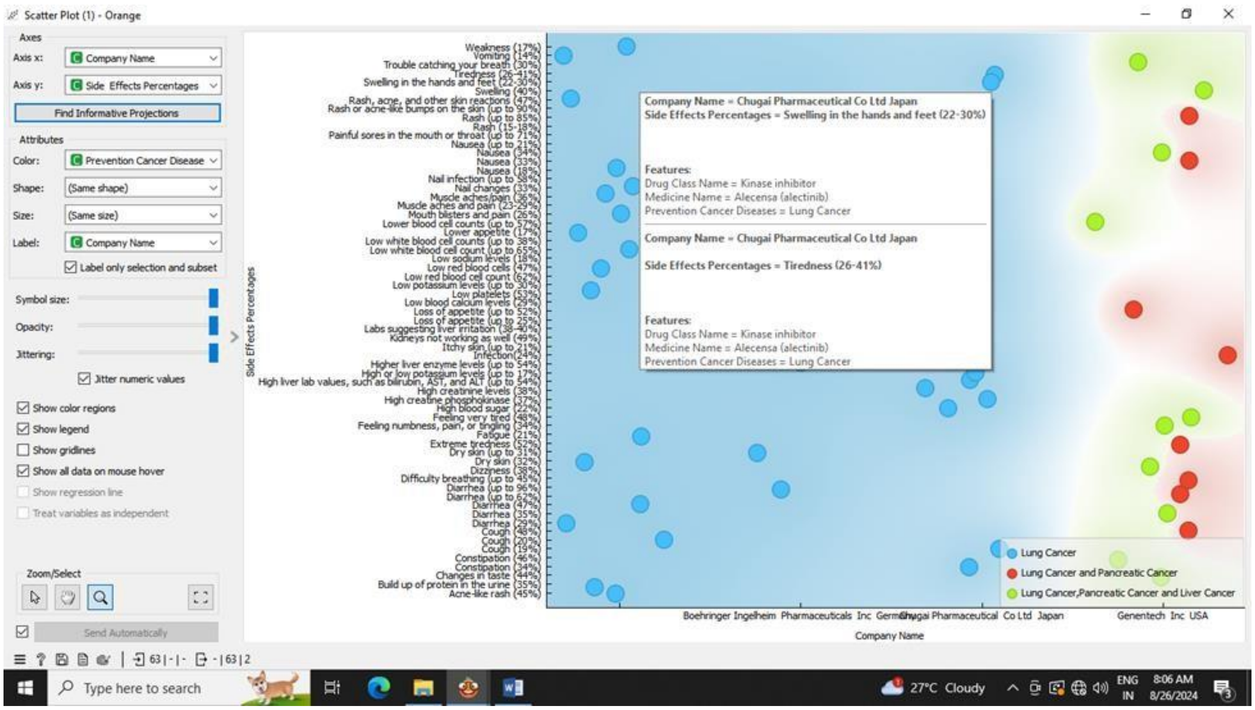
Graph 6: Drugs Class Name vs Side Effects Percentages

The graph 6 represented with visually x axis on Drug Class Name and y axis on Side Effects Percentages, these interacted with the scatter plot to explore multiple datasets relationship simultaneously displayed below represented a structured formats.

Drug Class Name = Tyrosine kinase inhibitor
Side Effects Percentage = Lower blood cell counts (up to 57%)
Feature
Company Name = Genentech Inc USA
Medicine Name = Rozlytrek (Entrectinib)
Prevention Cancer Diseases = Lung Cancer, Pancreatic Cancer and Liver Cancer

f. Company Name vs Side Effects Percentages:

To view and analysis the relationship between a company name and side effects percentage using a scatter plot with the x-axis and y-axis.



Graph 7: Company Name vs Side Effects Percentages

The graph 7 represented with visually same company two different medicines datasets [1] x axis on Company name and y axis on Side Effects Percentages [2]. X axis on Company name and y axis on Side Effects Percentage, these interacted with the scatter plot to explore multiple datasets relationship simultaneously displayed below represented a structured formats.

Company Name = Chugai Pharmaceutical Co Ltd Japan
Side Effects Percentage = Swelling in the hands and feet (22-30%)
Features
Drug Class Name = Kinase inhibitor
Medicine Name = Alecensa (alectinib)
Prevention Cancer Diseases = Lung Cancer

Company Name = Chugai Pharmaceutical Co Ltd Japan
Side Effects Percentage = Tiredness (26-41%)
Features
Drug Class Name = Kinase inhibitor
Medicine Name = Alecensa (alectinib)
Prevention Cancer Diseases = Lung Cancer

Scatter plot in Orange that visualizes multiple variables simultaneously is a great way to explore data, including handling both numerical and categorical dataset's. Scatter plot in Orange using multiple variables from a dataset by leveraging different encodings (color, size), user can be effectively visualize and analyze complex relationships in large datasets.

12. CONCLUSION AND FUTURE RESEARCH WORK:

The data mining Orange software has comprehensive exploration of multidimensional data through its graphical representation and integrated analysis tools. Scatter plots visualization has ability to handle complex datasets and provide intuitive visualization makes it a powerful tool for researcher, analysis and data scientists aiming to uncover insights from multidimensional data sources. Researchers can explore various data mining and machine learning techniques through Orange's extensive library of widgets. Orange software's scatter plot capabilities are indispensable for visualizing and analyzing relationships across diverse domains, empowering researchers to derive meaningful insights from their data and make informed decisions in their respective fields of future study. In future research direction using with scatter plot visualization techniques to find out the correlation ship between body mass index and cholesterol levels in cardiovascular diseases patients.

REFERENCES:

[1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians,* 74(3), 229-263. Google Scholar↗

[2] Sarkar, R. D. (2023). Pharma Software – A Complete Overview. *International Journal of Science and Healthcare Research,* 8(2). 166-177.Google Scholar↗

[3] Kiriiri, G. K., Njogu, P. M., & Mwangi, A. N. (2020). Exploring different approaches to improve the success of drug discovery and development projects: a review. *Future Journal of Pharmaceutical Sciences,* 6(27), 1-12. Google Scholar↗

[4] Nooreldeen, R., & Bach, H. (2021). Current and future development in lung cancer diagnosis. *International journal of molecular sciences*, 22(16), 8661. Google Scholar↗

[5] Chen, L., Chen, W. D., Xu, Y. X., Ren, Y. Y., Zheng, C., Lin, Y. Y., & Zhou, J. L. (2024). Strategies for Enhancing Non-Small Cell Lung Cancer Treatment: Integrating Chinese Herbal Medicines with Epidermal Growth Factor Receptor-Tyrosine Kinase Inhibitors Therapy. *European Journal of Pharmacology*, 980(176871). 1-2 Google Scholar↗

[6] Sun, X., Xu, S., Yang, Z., Zheng, P., & Zhu, W. (2021). Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors for the treatment of non-small cell lung cancer: a patent review (2014-present). *Expert Opinion on Therapeutic Patents,* 31(3), 223-238. Google Scholar↗

[7] Liu, T. C., Jin, X., Wang, Y., & Wang, K. (2017). Role of epidermal growth factor receptor in lung cancer and targeted therapies. *American Journal of Cancer Research*, 7(2), 187. Google Scholar↗

[8] Chen, L., Zhou, Y., Gan, C., Wang, X., Liu, Y., Dong, C. & Yang, J. (2022). Three third- generation epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer: Similarities and differences. *Cancer Investigation*, 40(7), 590-603. Google Scholar↗

[9] He, J., Huang, Z., Han, L., Gong, Y., & Xie, C. (2021). Mechanisms and management of 3rd-generation EGFR-TKI resistance in advanced non-small cell lung cancer. *International journal of*

oncology, 59(5), 1- 20. Google Scholar↗

[10] Araghi, M., Mannani, R., Heidarnejad maleki, A., Hamidi, A., Rostami, S., Safa, S. H., ... & Akhavan-Sigari, R. (2023). Recent advances in non-small cell lung cancer targeted therapy; an update review. *Cancer Cell International*, 23(1), 162. Google Scholar↗

[11] Han, J., Liu, Y., Yang, S., Wu, X., Li, H., & Wang, Q. (2021). MEK inhibitors for the treatment of non-small cell lung cancer. *Journal of Hematology & Oncology*, 14, 1-12. Google Scholar↗

[12] Niu, Z. X., Wang, Y. T., Lu, N., Sun, J. F., Nie, P., & Herdewijn, P. (2023). Advances of clinically approved small-molecule drugs for the treatment of non-small cell lung cancer. *European Journal of Medicinal Chemistry*,261(115868).2-3. Google Scholar↗

[13] Buszka, K., Ntzifa, A., Owecka, B., Kamińska, P., Kolecka-Bednarczyk, A., Zabel, M., & Budna-Tukan, J. (2022). Liquid biopsy analysis as a tool for TKI-based treatment in non-small cell lung cancer. *Cells*, 11(18), 2871. Google Scholar↗

[14] Savoska, S., & Ristevski, B. (2020). Towards implementation of big data concepts in a l pharmaceutica company. *Open Computer Science*, 10(1), 343-356. Google Scholar↗

[15] Bosilj-Vukšić, V., & Spremić, M. (2005). ERP System Implementation and Business Process Change: Case study of a pharmaceutical company. *Journal of computing and Information Technology*, 13(1), 11- 24. Google Scholar↗

[16] Subramanian, G., Thampy, A. S., Ugwuoke, N. V., & Ramnani, B. (2021). Crypto pharmacy–digital medicine: A mobile application integrated with hybrid blockchain to tackle the issues in pharma supply chain. *IEEE Open Journal of the Computer Society*, 2(3049330), 26-37. Google Scholar↗

[17] Matallah, H., Belalem, G., & Bouamrane, K. (2021). Comparative study between the MySQL relational database and the MongoDB NoSQL database. *International Journal of Software Science and Computational Intelligence* (IJSSCI), 13(3), 38-63. Google Scholar↗

[18] Ali, A., Naeem, S., Anam, S., & Ahmed, M. M. (2023). A state of art survey for big data Processing and nosql database architecture. *International Journal of Computing and Digital Systems*, 14(1), 1-1 Google Scholar↗

[19] Molke, A., Bhagat, R., & Gahat, V. (2024). Empowering Insights: The Power of Data Visualization with Power BI. *SSGM Journal of Science and Engineering*, 2(1), 77-81.Google Scholar↗

[20] Huang, Y., Wu, R., He, J., & Xiang, Y. (2024). Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: A comparative analysis with SAS, SPSS, and R. *Journal of Global Health*, 14. (04070), 5-6. Google Scholar↗

[21] Ogunleye, G. O., Fashoto, S. G., Daramola, C. Y., Ogundele, L. A., & Ojewumi, T. O. (2019). Development of a simple graphical interface based software for machine learning and data. visualization *International Journal of Recent Technology and Engineering*, 8(2), 3770-3777. Google Scholar↗

[22] Arief, I., Hasan, A., Putri, N. T., & Rahman, H. (2023). Literature Reviews of RBV and KBV Theories Reimagined-A Technological Approach Using Text Analysis and Power BI Visualization. JOIV: *International Journal on Informatics Visualization*, 7(4), 2532-2542. Google Scholar↗

[23] Kumar, A., Tejaswini, P., Nayak, O., Kujur, A. D., Gupta, R., Rajanand, A., & Sahu, M. (2022, May). A survey on IBM watson and its services. *In Journal of Physics: Conference Series* .2273(1), 1742-6596 IOP Publishing. Google Scholar↗

[24] Harfoushi, O., & Hasan, D. (2018). Amazon Machine Learning vs. Microsoft Azure Machine Learning as Platforms for Sentiment Analysis. *International Journal of Advanced Science and Technology,* 118(120), 131- 142. Google Scholar↗

[25] Sakai, K., Nagata, T., Mori, T., Inoue, S., Fujiwara, H., Odagami, K., ... & Mori, K. (2024). Research topics in occupational medicine, 1990–2022: A text-mining-applied bibliometric study. *Scandinavian Journal of Work, Environment & Health*, 50(7), 567-576. Google Scholar↗

[26] Ratra, R., & Gulia, P. (2020). Experimental evaluation of open source data mining tools (WEKA and Orange). *International Journal of Engineering Trends and Technology*, 68(8), 30-35. Google Scholar↗

[27] Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source, Software*, 6(60),2-3. Google Scholar↗

[28] Lone, H., & Warale, P. (2022). Cluster Analysis: Application of K-Means and Agglomerative Clustering for Customer Segmentation. *Journal of Positive School Psychology*. 6(5), 7798-7804. Google Scholar↗

[29] Sonkar, J., Maney, P., Yu, Q., & Palaiologou, A. (2019). Retrospective study to identify associations between clinician training and dental implant outcome and to compare the use of MATLAB with SAS. *International Journal of Implant Dentistry*, 5(28), 1-6. Google Scholar↗

[30] Raihen, M. N., Begum, S., Akter, S., & Sardar, M. N. (2025). Leveraging Data Mining for Inference and Prediction in Lung Cancer Research. *Computational Journal of Mathematical and Statistical Sciences*, 4(1), 139-161. Google Scholar↗

[31] Dobesova, Z. (2024). Evaluation of Orange data mining software and examples for lecturing machine learning tasks in geoinformatics. *Computer Applications in Engineering Education*,32(22735). 5-18. Google Scholar↗

[32] Hot, A., Mählitz, M., Wien-Nicolini, P., & Stegmüller, R. (2019). AstraZeneca versus Pfizer. *European Journal of International Management*, 13(5), 637-661. Google Scholar↗

[33] Ogura, T., Inoue, Y., Azuma, A., Homma, S., Kondoh, Y., Tanaka, K., ... & Nukiwa, T. (2023). Real-world safety and tolerability of nintedanib in patients with idiopathic pulmonary fibrosis: interim

report of a post-marketing surveillance in Japan. *Advances in Therapy*, 40(4), 1474-1493. Google Scholar↗

[34] Gelatti, A. C., Drilon, A., & Santini, F. C. (2019). Optimizing the sequencing of tyrosine kinase inhibitors (TKIs) in epidermal growth factor receptor (EGFR) mutation-positive non-small cell lung cancer (NSCLC). *Lung cancer*, 137(10), 113-122. Google Scholar↗

[35] Jassim, M. M., & Jaber, M. M. (2022). Systematic review for lung cancer detection and lung nodule classification: Taxonomy, challenges, and recommendation future works. *Journal of Intelligent Systems*, 31(1), 944-964. Google Scholar↗

[36] Honeywell, R. J., Kathmann, I., Giovannetti, E., Tibaldi, C., Smit, E. F., Rovithi, M. N., ... & Peters, G. J. (2020). Epithelial transfer of the tyrosine kinase inhibitors erlotinib, gefitinib, afatinib, crizotinib, sorafenib, sunitinib, and dasatinib: implications for clinical resistance. *Cancers*, 12(11), 3322. Google Scholar↗

[37] Nooreldeen, R., & Bach, H. (2021). Current and future development in lung cancer diagnosis. *International journal of molecular sciences*, 22(16), 8661. Google Scholar↗

[38] Onoi, K., Chihara, Y., Uchino, J., Shimamoto, T., Morimoto, Y., Iwasaku, M., ... & Takayama, K. (2020). Immune checkpoint inhibitors for lung cancer treatment: a review. *Journal of clinical medicine*, 9(5), 1362. Google Scholar↗

[39] Zugazagoitia, J., & Paz-Ares, L. (2022). Extensive-stage small-cell lung cancer: first-line and second-line treatment options. *Journal of Clinical Oncology*, 40(6), 671-680. Google Scholar↗

[40] Daly, M. E., Singh, N., Ismaila, N., Antonoff, M. B., Arenberg, D. A., Bradley, J., ... & Simone, C. B. (2022). Management of stage III non–small-cell lung cancer: ASCO guideline. *Journal of Clinical Oncology*, 40(12), 1356-1384. Google Scholar↗

[41] Petrella, F., Rizzo, S., Attili, I., Passaro, A., Zilli, T., Martucci, F., ... & Spaggiari, L. (2023). Stage III non-small-cell lung cancer: an overview of treatment options. *Current oncology*, 30(3), 3160-3175. Google Scholar↗

[42] Ungkulpasvich, U., Hatakeyama, H., Hirotsu, T., & di Luccio, E. (2023). Pancreatic Cancer and Detection Methods. *Biomedicines*, 11(9), 2557. Google Scholar↗

[43] Cao, W., Chen, H. D., Yu, Y. W., Li, N., & Chen, W. Q. (2021). Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. *Chinese medical journal*, 134(7), 783-791. Google Scholar↗

[44] Siregar, H. A., Raditya, M. Z., Yesa, A. N., & Permana, I. (2023). Comparison of Classification Algorithm Performance for Diabetes Prediction Using Orange Data Mining. *Indonesian Journal of Data and Science*, 4(3), 176-182. Google Scholar↗

[45] Shrivastava, A., Jaın, J., & Chauhan, D. (2023). Literature review on tools & applications of data mining. *International Journal of Computer Sciences and Engineering*, 11(4), 46- 54.

Google Scholar↗

[46] Ranjan, J. (2009). Data mining in pharma sector: benefits. *International journal of health care quality assurance*, 22(1), 82-92. Google Scholar↗

[47] Gillespie, J. J., Privitera, G. J., & Gaspero, J. (2019). Biopharmaceutical entrepreneurship, open innovation, and the knowledge economy. *Journal of Innovation Management*, 7(2), 59-77. Google Scholar↗

[48] Odoh, S. I., Michael, N. B., & Oche, P. S. (2024) Development of a Knowledge-based Patient Self-assessment and Diagnostic System for Malaria, Typhoid, and Related Diseases using Knowledge Discovery Database Techniques. *International Journal of Computer Applications*, 186(24),40- 43 Google Scholar↗

[49] Qi, W., Zhu, X., He, D., Wang, B., Cao, S., Dong, C. & Kang, J. (2024). Mapping Knowledge Landscapes and Emerging Trends in AI for Dementia Biomarkers: Bibliometric and Visualization Analysis. *Journal of Medical Internet Research*, 26(1), 1- 4. Google Scholar↗

[50] Estupiñán Ricardo, J., Domínguez Menéndez, J. J., Barcos Arias, I. F., Macías Bermúdez, J. M., & Moreno Lemus, N. (2021). Neutrosophic K-means for the analysis of earthquake data in Ecuador. *Neutrosophic Sets and Systems*, 44(1), 29. Google Scholar↗

[51] Novianto, R., & Triraharjo, B. (2024). Implementation of Orange Data Mining to Predict Student Graduation on Time at Pringsewu Muhammadiyah University. *Buana Information Technology and Computer Sciences (BIT and CS)*, 5(1), 29-38. Google Scholar↗

*****