data

MDPI

Article

VISEMURE: A Visual Analytics System for Making Sense of Multimorbidity Using Electronic Medical Record Data

Maede S. Nouri [1], Daniel J. Lizotte [2], Kamran Sedig [1,*] and Sheikh S. Abdullah [1]

[1] Insight Lab, Western University, London, ON N6A 3K7, Canada; mnouri4@uwo.ca (M.S.N.); sabdul9@uwo.ca (S.S.A.)
[2] Department of Computer Science, Faculty of Science and Department of Epidemiology and Biostatistics, Western University, London, ON N6A 3K7, Canada; dlizotte@uwo.ca
* Correspondence: sedig@uwo.ca; Tel.: +1-519-661-2111

Abstract: Multimorbidity is a growing healthcare problem, especially for aging populations. Traditional single disease-centric approaches are not suitable for multimorbidity, and a holistic framework is required for health research and for enhancing patient care. Patterns of multimorbidity within populations are complex and difficult to communicate with static visualization techniques such as tables and charts. We designed a visual analytics system called VISEMURE that facilitates making sense of data collected from patients with multimorbidity. With VISEMURE, users can interactively create different subsets of electronic medical record data to investigate multimorbidity within different subsets of patients with pre-existing chronic diseases. It also allows the creation of groups of patients based on age, gender, and socioeconomic status for investigation. VISEMURE can use a range of statistical and machine learning techniques and can integrate them seamlessly to compute prevalence and correlation estimates for selected diseases. It presents results using interactive visualizations to help healthcare researchers in making sense of multimorbidity. Using a case study, we demonstrate how VISEMURE can be used to explore the high-dimensional joint distribution of random variables that describes the multimorbidity present in a patient population.

Keywords: multimorbidity; visual analytics; conditional probability; binary logistic regression; softmax regression; decision tree; electronic medical record data

Citation: Nouri, M.S.; Lizotte, D.J.; Sedig, K.; Abdullah, S.S. VISEMURE: A Visual Analytics System for Making Sense of Multimorbidity Using Electronic Medical Record Data. *Data* **2021**, *6*, 85. https://doi.org/10.3390/data6080085

Academic Editor: Rüdiger Pryss

Received: 6 July 2021
Accepted: 30 July 2021
Published: 4 August 2021

1. Introduction

Multimorbidity, which refers to the presence of more than one chronic condition in a patient [1], is a prominent problem in healthcare. It is more prevalent in elderly patients, and is associated with higher morbidity, mortality, and increased healthcare costs [2]. Patients suffering from multiple chronic conditions are usually high-need and high-cost patients [3]. According to a study in 2015, the prevalence of multimorbidity was above 75 percent among elderly patients, and the total cost related to multimorbidity was 5.5 times higher than for other patients [4]. The higher the number of coexisting conditions a patient has, the more challenging it becomes to manage their care [5,6]. In 2012, approximately 38 million deaths worldwide were related to multiple chronic diseases; according to the World Health Organization, this number will increase to 52 million by 2030 [7].

Thus, there is a rising demand for research that provides deeper insights into multimorbidity. Electronic medical records (EMRs) hold great promise to facilitate the understanding of problems related to multimorbidity and its underlying mechanisms [8]. EMRs contain patient data such as prescriptions, demographics, diagnosis history, laboratory test results, discharge summaries, and surgical notes [9]. EMR databases are systematized platforms that can help medical professionals to access accurate and complete information about patients. With the progression of information technology and the extensive use of computerized systems, EMRs are available nowadays for subsequent use for research purposes [10–13]. For instance, EMRs can potentially aid clinical researchers in detecting

hidden patterns and trends, revealing missing events, identifying event sequences, establishing quality control, and reducing medical errors [14,15]. A number of recent studies have used EMRs to address multifaceted challenges of multimorbidity [16–18].

Despite the advantages of EMRs, it is often challenging for medical professionals to keep pace with the large quantity of heterogeneous data stored in EMRs [19]. These databases are usually complex and difficult to analyze and interpret. Automated data analysis methods based on statistics, data mining, and machine learning have the potential to fulfill the computational demands of EMRs [20,21]. Data analysis refers to the analysis of raw data to gain both deeper and novel insights into associations among the data elements [22]. However, one of the challenges of using such analysis methods lies in their lack of interpretability and transparency, which limits their application in EMR-based systems [12,13]. In order to overcome this challenge, it is possible to make the analysis processes accessible to the user through interactive visualizations. A new set of computational tools, known as visual analytics systems (VASes), have the potential to help reduce the complexity of EMRs by combining automated analysis techniques with interactive visualizations [23–25]. VASes can help with the analysis, interpretation, and making sense of EMR databases by improving the capabilities of the user to accomplish complex data-driven tasks [26]. Even though VASes hold great promise for analyzing and making sense of EMR data from patients with multiple chronic diseases, up until now, there is a shortage of VASes for understanding multimorbidity.

The objective of this study is to demonstrate how VASes can be designed to offer deeper insights into multimorbidity by enabling exploratory analyses of EMR datasets and providing a rich set of descriptive statistics. We provide a foundation for, and implementation of, the design and use of VA systems in exploring the prevalence and patterns of multimorbidity in a given patient population. To this end, we present a novel web-based system that we have developed, called VISEMURE—VIsual analytics system for making SEnse of MUltimorbidity using electronic medical REcord. To illustrate the usefulness of VISEMURE, we use the Deliver Primary Healthcare Information project database, which is available through the Canadian Primary Care Sentinel Surveillance Network [27].

The proposed system uses an interactive bar chart to display the prevalence of chronic diseases, as well as a dynamic correlation matrix to present the correlations among occurrences of those diseases. Disease prevalence and correlations may be estimated using count-based conditional probability, logistic regression, and decision tree models. VISEMURE can also create conditional prevalence and correlation estimates, based on any pre-existing conditions the patient may have, and on other patient characteristics such as gender, age, household income, and household education. This allows for the investigation of the impact of existing chronic disease and patient characteristics on the distribution of multimorbidity in a patient population. The visualization techniques in VISEMURE can be repurposed for other tasks in the area of healthcare where high-dimensional joint distributions of random variables are important to understand.

We envision VISEMURE being most useful for researchers who are familiar with EMR data and for multidisciplinary teams of data specialists and clinical specialists to investigate many different questions related to multimorbidity. Because multimorbidity is a complex phenomenon that is not easily captured by a small number of prespecified visualizations, we contend that a VAS is needed. The main purpose of VISEMURE is not to create plots for a static publication (although it could be used for this task) but rather to quickly and efficiently make sense of the multimorbidity patterns within a dataset and to generate hypotheses about how these patterns may generalize to other settings.

The rest of this paper is organized as follows: Section 2 explains the methodology employed for VISEMURE. Section 3 explains the design of the proposed system by describing its structure and components. Section 4 presents the results using some case studies to illustrate the usefulness of the system. Sections 5 and 6 include discussion, limitations, and some future areas of application. Finally, Section 7 presents the conclusion of the paper.

2. Methods

This section describes the methodology we have employed to design the proposed VA system, namely VISEMURE. In Section 2.1, we explain the data source. We then describe the preprocessing steps in Section 2.2. Next, in Section 2.3, we introduce the analytical and visual components of VISEMURE and briefly describe how these components are combined, which is discussed more extensively in Section 3. Finally, Section 2.4 outlines the implementation details of VISEMURE.

2.1. Data Source

VISEMURE is designed to be used with *structured* EMR data, that is, data that are in a tabular format consisting of columns with well-defined entries for the variables of interest, including patient characteristics such as age and sex, and disease status. Many EMR data sources have this structure, but if a data source does not, applying case definitions [27] and/or natural language processing [28] to create patient characteristics or disease status variables may be required as a preprocessing step.

To demonstrate the use of VISEMURE, we use a subset of the Deliver Primary Healthcare Information (DELPHI) project database. It is one of the eleven regional networks included in the Canadian Primary Care Sentinel Surveillance Network [29]. DELPHI established the first Canadian primary care database derived from EMR data, which coded symptoms and diagnoses for a subset of patient encounters using the International Classification of Primary Care.

For our illustrative example, we use a subset of the DELPHI database that includes a total of 13,697 patients who have at least one of 20 specified chronic diseases. Each patient is further characterized by three features: age, gender, and socioeconomic status (SES). Among a total of 7565 females and 6132 males in the dataset, 6303 patients have developed only one disease, 3183 patients have developed two chronic diseases, and 4211 patients have developed more than two chronic conditions. SES is categorized into five equal-sized quintiles. The first quintile represents the lowest-income group, whereas the fifth quintile refers to the highest-income group. The distribution of sociodemographic factors among 13,697 patients is shown in Table 1. Small cell sizes have been suppressed. It is important to note that there are no patients in the first and second quintile for both female and male categories; hence, we would not be able to use these data to extrapolate conclusions to patients in these income quintiles.

Table 1. The distribution of sociodemographic factors among the patients (i.e., frequency in each category).

	Female			**Male**			**Total**
	Third Quintile	**Fourth Quintile**	**Fifth Quintile**	**Third Quintile**	**Fourth Quintile**	**Fifth Quintile**	
Child	<5	58	20	<5	83	17	178
Percentages	<3%	32.58%	11.23%	<3%	46.63%	9.55%	100%
Adolescent	<5	154	33	<5	155	47	391
Percentages	<2%	39.39%	8.44%	0%	39.64%	12.02%	100%
Young Adult	7	312	142	6	210	58	735
Percentages	0.95%	42.45%	19.32%	0.81%	28.57%	7.89%	100%
Adult	8	493	157	6	340	86	1090
Percentages	0.73%	45.23%	14.40%	0.55%	31.19%	7.89%	100%
Middle Age	20	2077	647	29	1678	491	4942
Percentages	0.40%	42.03%	13.09%	0.59%	33.95%	9.93%	100%
Elder	9	2423	1003	18	2106	802	6361
Percentages	0.14%	38.09%	15.77%	0.28%	33.11%	12.61%	100%
Total	46	5517	2002	59	4572	1501	13,697
Percentages	0.33%	40.28%	14.62%	0.43%	33.38%	10.96%	100%

Table 2 depicts the list of twenty chronic diseases ordered by patient counts according to the dataset, which was derived from the DELPHI database using the same methodology as Nicholson (2017) [30]. As shown in Table 2, 'Hypertension', 'Hyperlipidemia', and 'Bronchitis' are the most common diseases, whereas 'Kidney Disease', 'Dementia', and 'Liver Disease' are the least common diseases among all patients in the database.

Table 2. The distribution of chronic diseases among the patients.

	Chronic Disease	Patient Counts	%
1	Hypertension	4345	31.72
2	Hyperlipidemia	3442	25.13
3	Bronchitis	2617	19.11
4	Cardiovascular Disease	2332	17.02
5	Musculoskeletal Problem	2163	15.79
6	Diabetes	2161	15.78
7	Depression	1747	12.75
8	Arthritis	1718	12.54
9	Cancer	1589	11.60
10	Thyroid Disease	1510	11.02
11	Obesity	1266	9.24
12	Colon Problem	1216	8.88
13	Osteoporosis	926	6.76
14	Urinary Problem	861	6.29
15	Stomach Problem	804	5.87
16	Heart Failure	306	2.23
17	Stroke	231	1.69
18	Kidney Disease	212	1.55
19	Dementia	210	1.53
20	Liver Disease	45	0.33

We have chosen the ten most common chronic diseases based on our dataset to use for further exploratory analysis. They are as follows: 'Hypertension', 'Hyperlipidemia', 'Bronchitis', 'Cardiovascular Disease', 'Musculoskeletal Problem', 'Diabetes', 'Depression', 'Arthritis', 'Cancer', and 'Thyroid Disease'. The main reason for this choice is that the dataset is not large enough to allow a good estimation of disease prevalence and correlations when prevalence is very low.

2.2. Preprocessing

This section describes the preprocessing steps to prepare the data for the statistical and machine learning techniques in VISEMURE.

2.2.1. Creating Dummy Variables

All chronic diseases, as well as gender, are already binary variables taking values either 0 or 1 in the dataset. Age is, however, a categorical variable with more than two categories and was converted into dummy variables prior to regression.

2.2.2. Merging Categories with Few Observations

Since some dummy variables in the dataset have very few observations, the classification models in VISEMURE were unable to fit models properly and return null or not-a-number (NaN) values as coefficients. This results in NaNs as prevalence and correlation estimates. One solution for this problem is to merge the categories of predictors with a small number of patients. To do so, the three groups of 'Child', 'Adolescent', and 'Young Adult' have been merged and labeled as 'Child and Young Adult'. We also merge 'Adult' and 'Middle-Aged' into one category. Therefore, the modified age variable in the dataset has three categories—namely, 'Child and Young Adult', 'Adult and Middle-Aged', and 'Elder'.

Similarly, 'Third Income Quintile' and 'Fourth Income Quintile' have also been merged. According to Statistics Canada [31], Table 3 shows the average adjusted after-tax income that is divided into five quintiles in 2010.

Table 3. Average Adjusted After-Tax Income by five quintiles for the population in 2010.

Quintile	Average Adjusted After-Tax Income
Lowest income quintile	USD 16,000
Second income quintile	USD 28,000
Third income quintile	USD 38,500
Fourth income quintile	USD 50,600
Highest income quintile	USD 85,500

Statistics Canada's income grouping is used to label the new categories of SES after merging them. This attribute breaks down the patients into two groups of 'Less than or Equal to USD 50,600' and 'Greater than USD 50,600' average adjusted after-tax income.

2.3. Components of VISEMURE

This section introduces VISEMURE by providing an overview of its design components. The main interface of VISEMURE is presented in Figure 1. VISEMURE calculates conditional probabilities and performs logistic regression, softmax regression, and decision tree on the data in real-time based on user selections. These statistical and machine learning models are currently the available techniques employed by the system to analyze and interactively visualize the input data. However, the system is modular and so can be extended by incorporating additional methods.

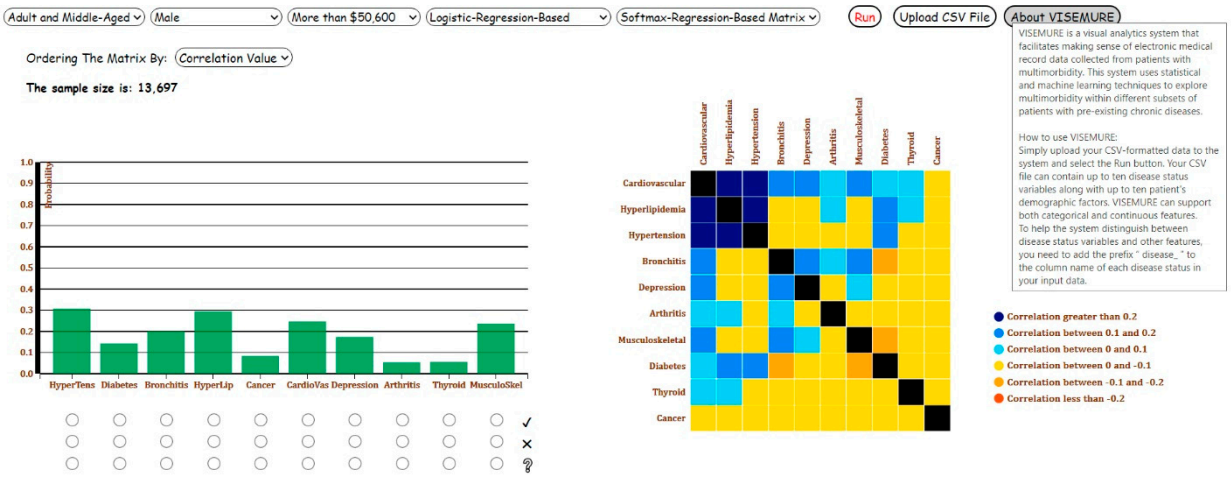

Figure 1. Overview of our proposed visual analytics system, VISEMURE.

First, five dropdown lists are created to allow the user to select different categories of age, gender, and SES available in the data, as well as the types of models applied to the visualizations in the system.

Second, a bar graph is designed interactively to represent the prevalence of chronic diseases. The three prevalence estimation techniques—namely, count-based conditional probability, logistic regression, and decision tree—are used by the bar graph. Based on the user selections, the results of the model of interest are shown in the bar chart. Three radio buttons (graphical control elements) are associated with each bar in the bar chart, allowing end-users to select only one item at a time. Every bar encodes the prevalence of the chronic disease, and its corresponding radio buttons take one of the labels among 1, 2, or null. If the user selects the first and nearest radio button to a bar with label 1, all prevalence estimates will be made conditional on having the corresponding disease. In contrast, selecting the second radio button produces prevalence estimates conditional on *not* having the corresponding disease. Finally, a radio button with a null label under each bar indicates that the prevalence estimates will be made assuming the status of that disease is unknown.

Third, a dynamic correlation matrix is created to show the pairwise correlations between chronic diseases. Two machine learning models, decision tree and softmax regression, are employed to estimate these correlation values. The user can select one of these two models from the corresponding dropdown menu. The system then computes the correlation coefficients based on the selected model. One additional dropdown menu is incorporated into the system to let the user order the cells in the correlation matrix by either disease name or correlation value.

When the data is filtered by the user, the sample size of the filtered data is shown in the interface. We note that VISEMURE is intended for exploratory analysis of a given dataset that represents a patient population, hence, there is no "minimum dataset size" required, subject to the needs of the different analysis techniques as described in Section 2.2.2. The development of VASes for analyses that generalize beyond the given dataset is an active area of research, and is discussed in Section 6.

VISEMURE can analyze and visualize arbitrary datasets containing up to ten features representing patient characteristics in addition to up to ten chronic diseases encoded as binary variables. We found that ten of each was sufficient to allow for a rich exploration of multimorbidity patterns, but this could be easily increased if a user had sufficient data and screen real estate. The system dynamically generates dropdown menus for features in input data, so it can be used with other data with other categorical and continuous patient characteristics.

An "About VISEMURE" button is provided such that when the user hovers over the button, a user guide appears that contains a brief introduction of the system as well as instructions on using it (see Figure 2).

2.4. Implementation Details

The VISEMURE system is designed using Flask and D3.js. Flask is a Python web application framework, and D3.js is a library in JavaScript for creating interactive visualizations. We built our binary logistic regression and softmax regression models with Python library Statsmodels [32] and decision tree model using python library Scikit-Learn [33].

We use D3.js to develop interactive visualizations primarily because D3 (1) provides a data-driven method to attach data to the DOM (i.e., Document Object Model) elements. (2) allows the user to access the full functionalities of state-of-the-art web-browsers, and (3) is compatible with other programming languages such as python.

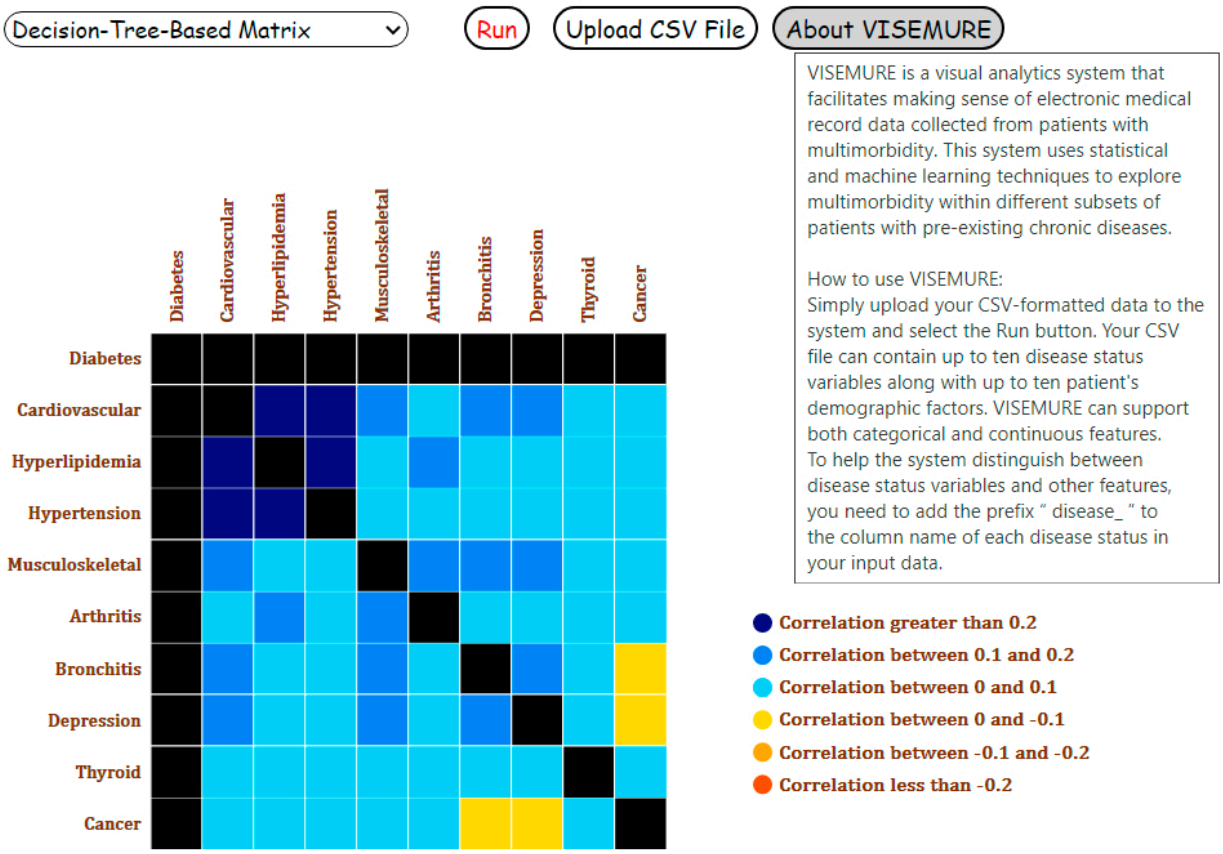

Figure 2. User guide of our proposed visual analytics system, VISEMURE.

2.5. Architecture

The architecture of VISEMURE is shown in Figure 3. VISEMURE is comprised of three modules: Analytics, Visualization, and Interaction. The Analytics module consists of two components: (1) disease prevalence estimator (shaped like an oval) and (2) disease correlations estimator (shaped like a rounded rectangle). The disease prevalence estimator can apply count-based conditional probability, logistic regression, or decision tree to compute prevalence estimates of chronic diseases. The visualization module displays these estimates through Count-Based Bar Chart, Logistic-Regression-Based Bar Chart, and Decision-Tree-Based Bar Chart, respectively. The disease correlations estimator employs a softmax regression or decision tree to produce correlation estimates between chronic diseases. The visualization module encodes the outputs of the disease correlations estimator into Softmax-Regression-Based Correlation Matrix and Decision-Tree-Based Correlation Matrix. The Interaction module of VISEMURE provides users with three main actions: (1) selecting, (2) filtering, and (3) arranging. Using the interaction module, users can gain insight into the data and explore associations between chronic conditions and patient characteristics by selecting drop-down menus and radio buttons. Users can also filter the data and display it in the Count-Based Bar Chart and observe the sample size of the filtered data or rearrange the correlation matrix to see the degree of association of diseases easily.

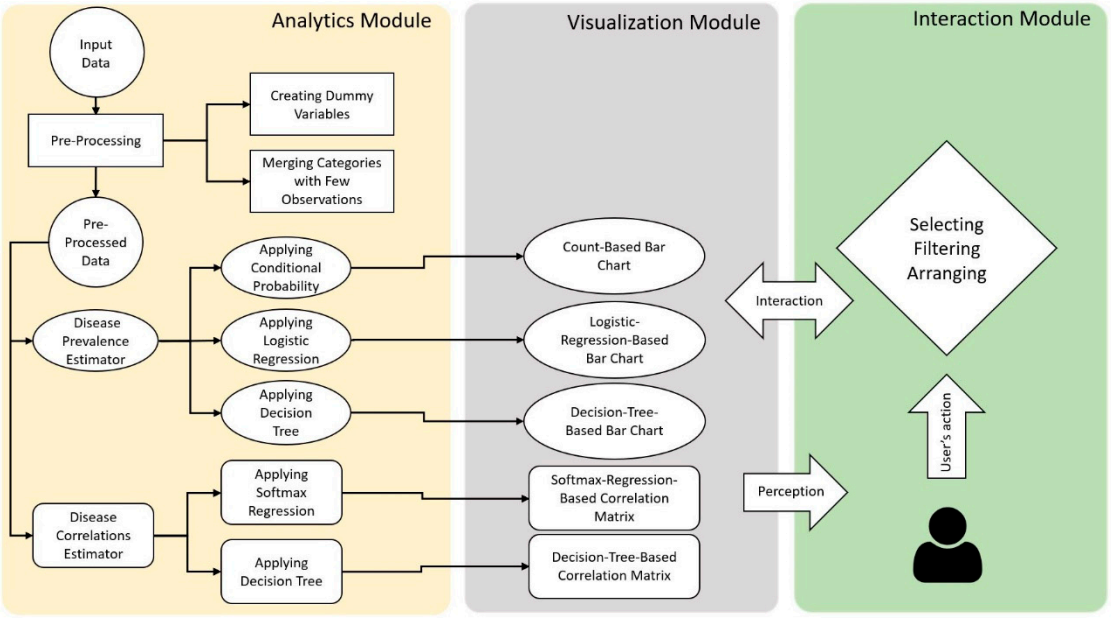

Figure 3. Architecture of our proposed visual analytics system, VISEMURE.

3. Design of VISEMURE

In this section, we explain the design of VISEMURE by describing how different visual and analytical components are combined in the system to facilitate various interactions. Section 3.1 describes how simple count-based estimates can be used by VISEMURE to produce and display prevalence and correlation estimates. Because such count-based estimates can be problematic when exploring small groups of patients, Sections 3.2 and 3.3 describe different machine learning techniques (i.e., court-, logistic regression-, and decision tree-based) that are also supported by the bar chart of VISEMURE. Next, Sections 3.4 and 3.5 describe how two other machine learning techniques (i.e., softmax regression- and decision tree-based) can be used to interactively generate the correlation matrix.

3.1. Count-Based Bar Chart

By selecting 'Count-Based Bar Chart' from the dropdown list corresponding to the type of the interactive bar chart, the prevalence of chronic diseases is displayed on the bar chart in our VA system. Each bar on the x-axis is allocated to one disease X_i; the prevalence of that disease $P(X_i = 1)$ is presented on the y-axis. If the user selects two diseases, the system calculates the probability of each unselected disease conditioned on the presence of both selected diseases. Then, the system animates the change and updates the visualization. The selection process can be continued by the user to look for further associations within the subgroup who have the selected diseases, and so on.

The user can also interact with the visualizations by selecting different age, gender, and socioeconomic groups from the dropdown lists. As a result, the dataset of multimorbid patients would be filtered based on the selected sociodemographic factors, and the conditional probabilities would be updated accordingly. For example, suppose the user selects 'Child and Young Adult' as the age group, 'Male' from the gender groups, and the existence of diabetes. The prevalence of each unselected disease would then be computed and presented on its associated bar, estimated only using patients who are child or young adult, male, with diabetes. Because it is selected and assumed to be present, the prevalence of diabetes would change to 1 in the bar graph. In this case, the conditional probability formula for the jth unselected disease is as follows

$$P\left(X_j = 1 \middle| \mathit{diabetes} = 1,\ \mathit{age} = \text{child and young adult},\ \mathit{gender} = \text{male}\right)$$

3.2. Logistic-Regression-Based Bar Chart

Selecting 'Logistic-Regression-Based Bar Chart' from the corresponding drop-down list, the bar graph represents the prevalence of chronic diseases predicted by logistic regression. The logistic regression model uses the entire data with all 13,697 patients and only the selected sociodemographic factor/s and pre-existing disease/s are included in these models. The model will be changed and updated if the user changes the selection. As an example, if the user clicks on the radio button with label 0 related to arthritis (the absence of arthritis), then selects the radio button with label 1 corresponding to thyroid disease (the presence of thyroid disease) and 'Elder' age group, the logistic regression model for finding a mathematical relationship between them and 'Cancer' as the target is as follows

$$z = \log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1(arthritis) + \beta_2(age1) + \beta_3(age2) + \beta_4(thyroid\ disease)$$

where P is the probability of developing cancer, which would be shown on the corresponding bar as the cancer prevalence. The model investigates the associations between arthritis and cancer, and thyroid disease and cancer, adjusted for the dummy variables *age1* and *age2*.

3.3. Decision-Tree-Based Bar Chart

When the user selects 'Decision-Tree-Based Bar Chart' from the dropdown list, a decision tree model will be created such that all diseases of interest, as well as selected patient characteristics, are included in the model. It is important to note that if a categorical variable with more than two categories is selected (e.g., age), we do not use one-hot encoding to binarize each category, which converts the categorical variable into dummy variables. We avoid this process because dummy variables make a decision tree sparse and obscure the order of feature importance, which results in inefficiency and poor performance. We build the model based on all patients included in the dataset. To avoid overfitting and to reduce complexity, we utilize pruning methods by changing the parameters 'max_depth' (=3) and 'min_samples_leaf' (=200) in the Python server, which refers to the maximum number of nodes in a branch and the minimum number of samples required at the leaf node (a node without further split), respectively.

3.4. Softmax-Regression-Correlation Matrix

Suppose we aim to measure the association between two chronic diseases D_1 and D_2. We create a new variable A having the following four levels:

$A = 0$ if $D_1 = 0$ and $D_2 = 0$ $A = 1$ if $D_1 = 0$ and $D_2 = 1$ $A = 2$ if $D_1 = 1$ and $D_2 = 0$ $A = 3$ if $D_1 = 1$ and $D_2 = 1$

Since our new target is the variable A with four levels (K = 4), we can build a softmax (or multiclass logistic) regression in order to predict the probabilities for each of the levels of A, which in turn can be used to compute the pairwise correlation between these two diseases. Softmax regression utilizes a linear predictor function $f(k,i)$ to predict the probability that observation i belongs to class k

$$f(k,i) = \beta_{0,k} + \beta_{1,k}x_{1,i} + \beta_{2,k}x_{2,i} + \ldots + \beta_{M,k}x_{M,i} \quad \text{for } k = 1, \ldots, K$$

where M is the number of independent variables in the model and i is an observation from 13,697 inputs in the data. We assign value 0 to the 'Male' category and value 1 to the 'Female' category, since in this dataset, gender is encoded as a binary variable. If the user selects the presence depression and 'Male' group, the softmax regression model built for class zero is as follows

$$f(0) = \beta_{0,0} + \beta_{1,0}(depression = 1) + \beta_{2,0}(gender = 0)$$

After computing the linear predictor function for all four classes of the dependent variable A, we can also compute the probability of each class as follows

$$P(A=0) = \frac{e^{\beta_{0,0}+\beta_{1,0}(depression=1)+\beta_{2,0}(gender=0)}}{1+\sum_{k=1}^{3} e^{f(k)}}$$
$$P(A=1) = \frac{e^{\beta_{0,1}+\beta_{1,1}(depression=1)+\beta_{2,1}(gender=0)}}{1+\sum_{k=1}^{3} e^{f(k)}}$$
$$P(A=2) = \frac{e^{\beta_{0,2}+\beta_{1,2}(depression=1)+\beta_{2,2}(gender=0)}}{1+\sum_{k=1}^{3} e^{f(k)}}$$
$$P(A=3) = \frac{1}{1+\sum_{k=1}^{3} e^{f(k)}}$$

The correlation between the two random variables X and Y is calculated through the following formula

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}$$

where σ_X and $E(X)$ denote the standard deviation and the expected value of X, respectively, and $E(XY)$ is defined as follows when X and Y are discrete random variables and not independent

$$E(XY) = \sum_{x \in X} \sum_{y \in Y} xy \; P(X=x, Y=y)$$

We name $P(A=0) = P_{00}$, $P(A=1) = P_{01}$, $P(A=2) = P_{10}$, and $P(A=3) = P_{11}$. We also define $P_{1.} = P_{10} + P_{11}$ and $P_{.1} = P_{01} + P_{11}$. Given that each chronic disease in our data can be modeled as a random variable with a Bernoulli distribution, we have $E(D_1) = P_{1.}$, $\sigma^2_{D_1} = P_{1.}(1 - P_{1.})$, $E(D_2) = P_{.1}$, and $\sigma^2_{D_2} = P_{.1}(1 - P_{.1})$.

According to the definition of $E(XY)$ and given that D_1 and D_2 might influence each other, we calculate $E(D_1 D_2) = (0 \times 0 \times P_{00}) + (0 \times 1 \times P_{01}) + (1 \times 0 \times P_{10}) + (1 \times 1 \times P_{11}) = P_{11}$. Then, the correlation etween D_1 and D_2 is computed as follows

$$\rho_{D1,D2} = \frac{P_{11} - P_{1.}P_{.1}}{\sqrt{P_{1.}(1 - P_{1.})}\sqrt{P_{.1}(1 - P_{.1})}}$$

This process is repeated for each pair of chronic diseases, and their estimated correlation is depicted by the corresponding cell in the interactive matrix. By hovering over each cell, the corresponding correlation value appears. The direction of the relationships between diseases is encoded by color. Blue and orange are used for positive and negative correlations, respectively. In addition, color intensity encodes the magnitude of the correlation coefficients such that a darker color represents a greater absolute value. The user can also rearrange the correlation matrix by disease name and correlation value. Recall that the height of the bar corresponding to a selected disease changes to 1 or 0, based on the selection (assumed presence or absence). Similarly, if the user selects *disease i*, the color of all cells in row i and the column i in the correlation matrix change to black, which indicates that the correlations are undefined.

3.5. Decision-Tree-Based Correlation Matrix

By selecting 'Decision-Tree-Based-Correlation Matrix' from the dropdown menu related to the type of matrix, a decision tree is made given the selected variables and with the parameters 'max_depth' = 3 and 'min_samples_leaf' = 200 to prevent overfitting. The target in the correlation matrix is the variable A corresponding to a pair of chronic diseases and has four levels. For instance, suppose the user selects 'Adult and Middle-Aged' and the presence of hyperlipidemia and aims to observe their influence on the association between cardiovascular disease and hypertension as the target. Therefore, the model would examine the relationship between hyperlipidemia and the target controlling for age. Then, the probability of occurring for each class of the target would be estimated using one instance (in this case *age* = 'Adult and Middle-Aged' and *hyperlipidemia* = 1). The four computed probabilities would be used in estimating the correlation coefficient between

cardiovascular disease and hypertension. This analysis would be repeated for all other pairs of unselected diseases. The correlation of those pairs whose one or both diseases are assumed to be known to be present or absent is undefined. In this example, all correlations between hyperlipidemia and the other nine chronic diseases would be undefined and their relative cells in the correlation matrix would change to black.

4. Results

VISEMURE can be used in an iterative manner. In this section, we explain the process of using the system through three case studies (i.e., Analysis 1, 2, and 3) to make it easier for the reader to follow. There can be a large number of action sequences through which users can accomplish their tasks. In our VA system, many different multimorbidity patterns can be explored through user selection and filtering. We note that the VISEMURE system is intended for exploratory investigation of a specified patient population, rather than for extrapolating to new populations or confirming relationships among variables. Hence, VISEMURE does not provide any hypothesis testing or confidence measures (besides sample size).

4.1. Analysis 1

Assume the user aims to estimate the marginal probability (prevalence) of diseases for 'Child and Young Adult' category. The data would be filtered on the age group of interest, and the results would be displayed on the Count-Based Bar Chart. As shown in Figure 4, bronchitis and depression are the most prevalent diseases among 1304 children and young adults. Figure 5 shows the correlations between the ten diseases on Softmax-Regression-Based Correlation Matrix in Analysis 1. The user can observe the number of cells denoted by orange is more than the number of blue cells, though all values are between −0.3 and 0.3, indicating weak correlations between the diseases.

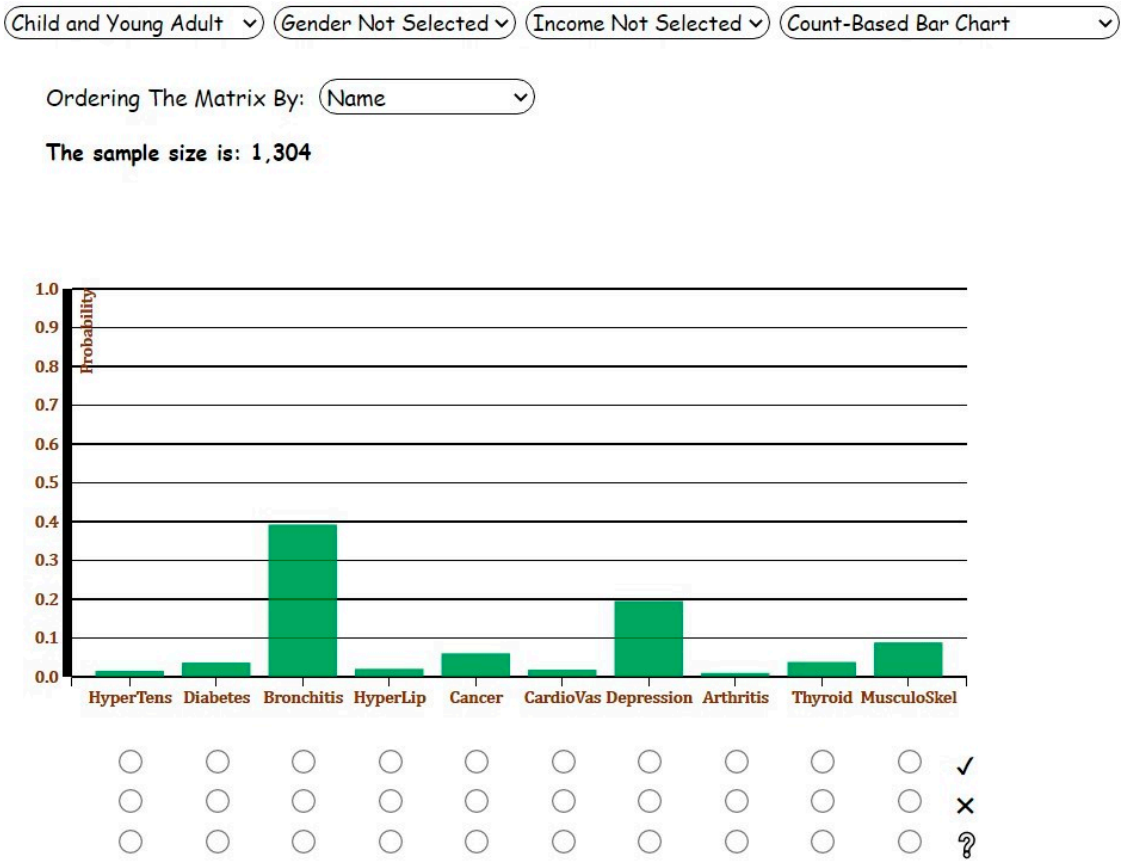

Figure 4. Screenshot of the Count-Based Bar Chart for Analysis 1 with 'Child and Young Adult' age category selected.

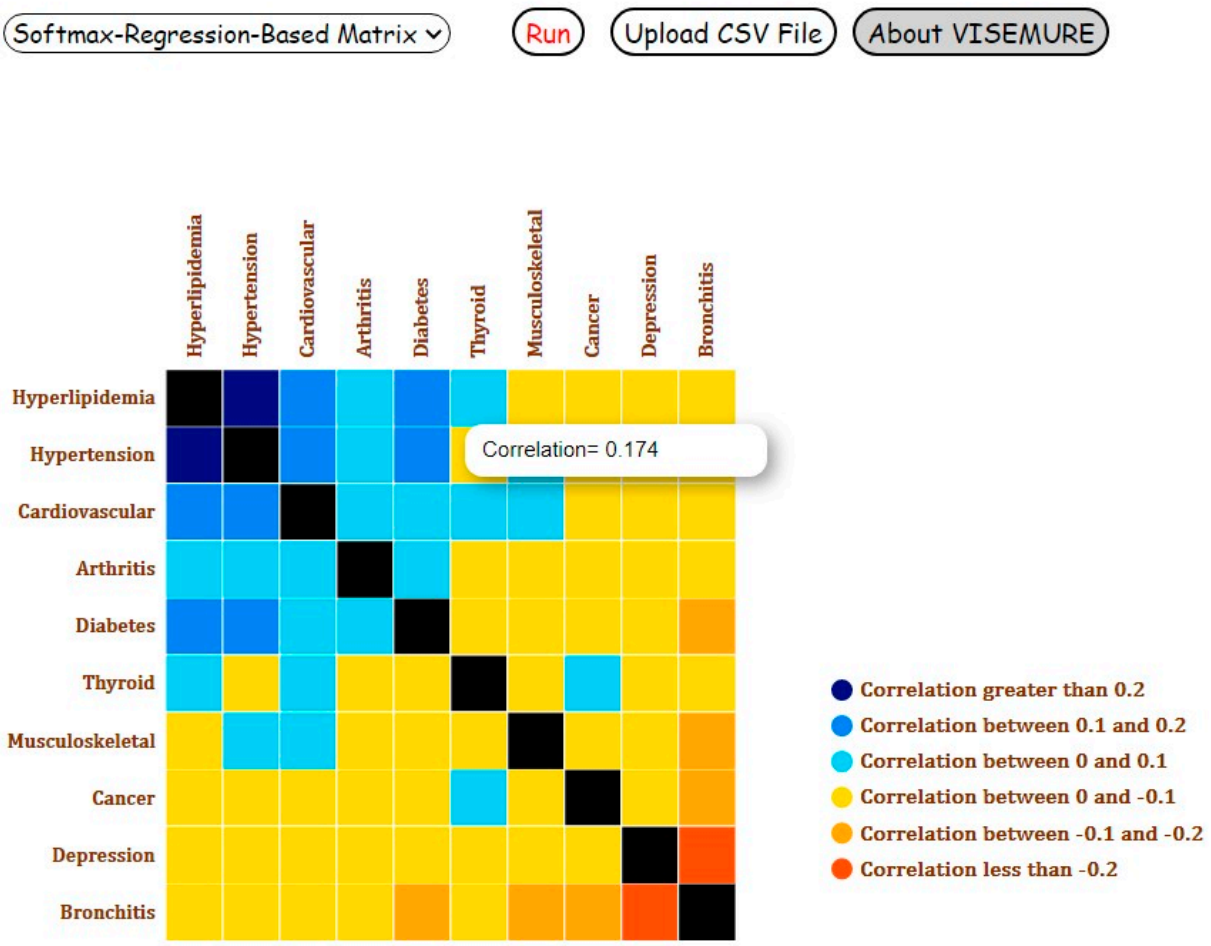

Figure 5. Screenshot of the Softmax-Regression-Based Correlation Matrix for Analysis 1 with 'Child and Young Adult' age category selected.

4.2. Analysis 2

Following this, suppose the user changes the age group to 'Elder', selects 'Female' category, and chooses Decision-Tree-Based Bar Chart from the fourth dropdown list to observe and interpret the results. Compared to Analysis 1, the prevalence of all chronic diseases increases noticeably except bronchitis and depression, which are more common among young adults and children (Figure 6).

4.3. Analysis 3

As the next step, suppose the user selects the radio buttons with label 1 for hypertension and arthritis. The probabilities of chronic diseases conditioned on the diagnosis of hypertension and arthritis would be represented on the corresponding bars. Depicted in Figure 7, in the presence of hypertension and arthritis, the prevalence of hyperlipidemia and musculoskeletal problem increases by seven percent and five percent, respectively. Similarly, Figure 8 shows the Decision-Tree-Based Correlation Matrix for Analysis 3 with 'Elder' and 'Female' categories selected.

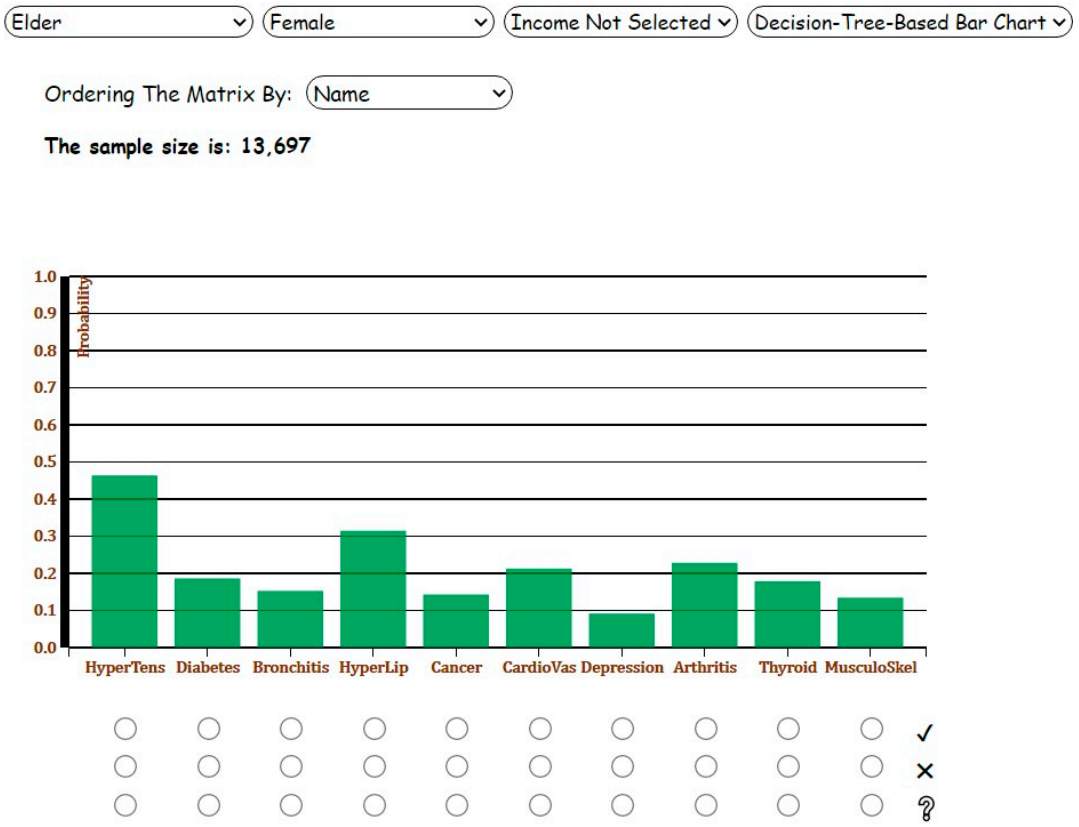

Figure 6. Screenshot of the Decision-Tree-Based Bar Chart for Analysis 2 with 'Elder' and 'Female' categories selected.

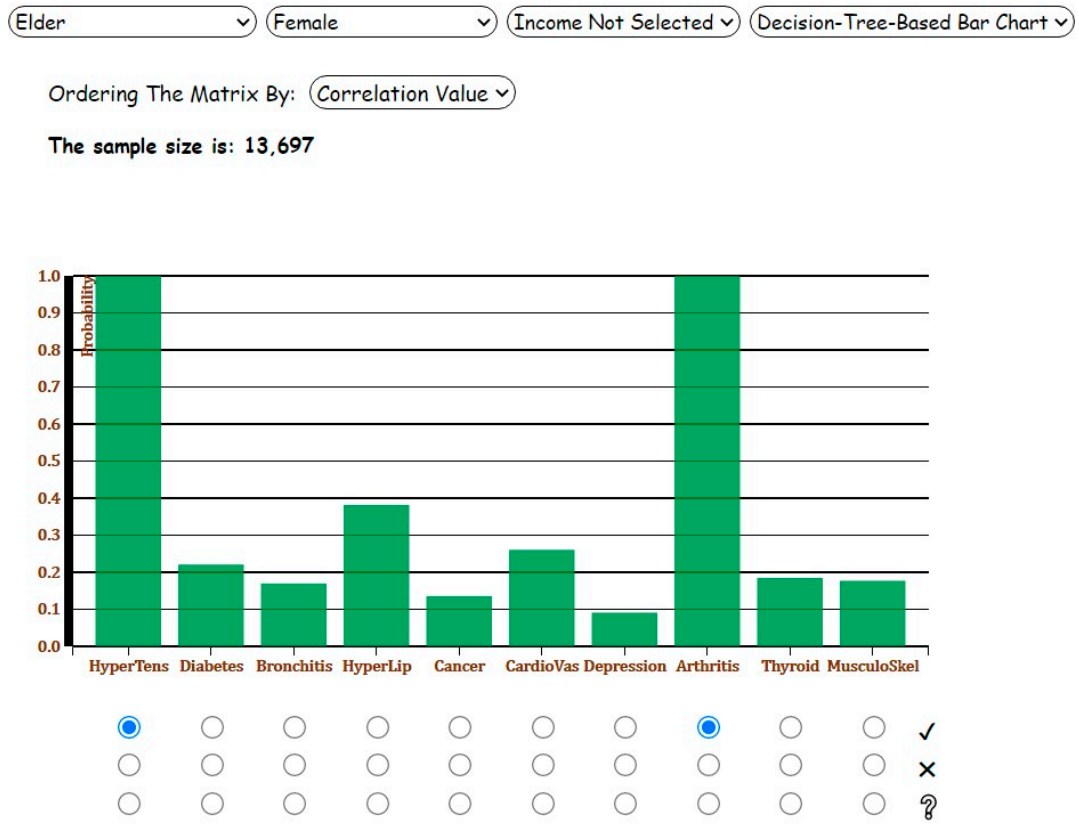

Figure 7. Screenshot of the Decision-Tree-Based Bar Chart for Analysis 3 with 'Elder' and 'Female' categories and the presence of hypertension and arthritis selected.

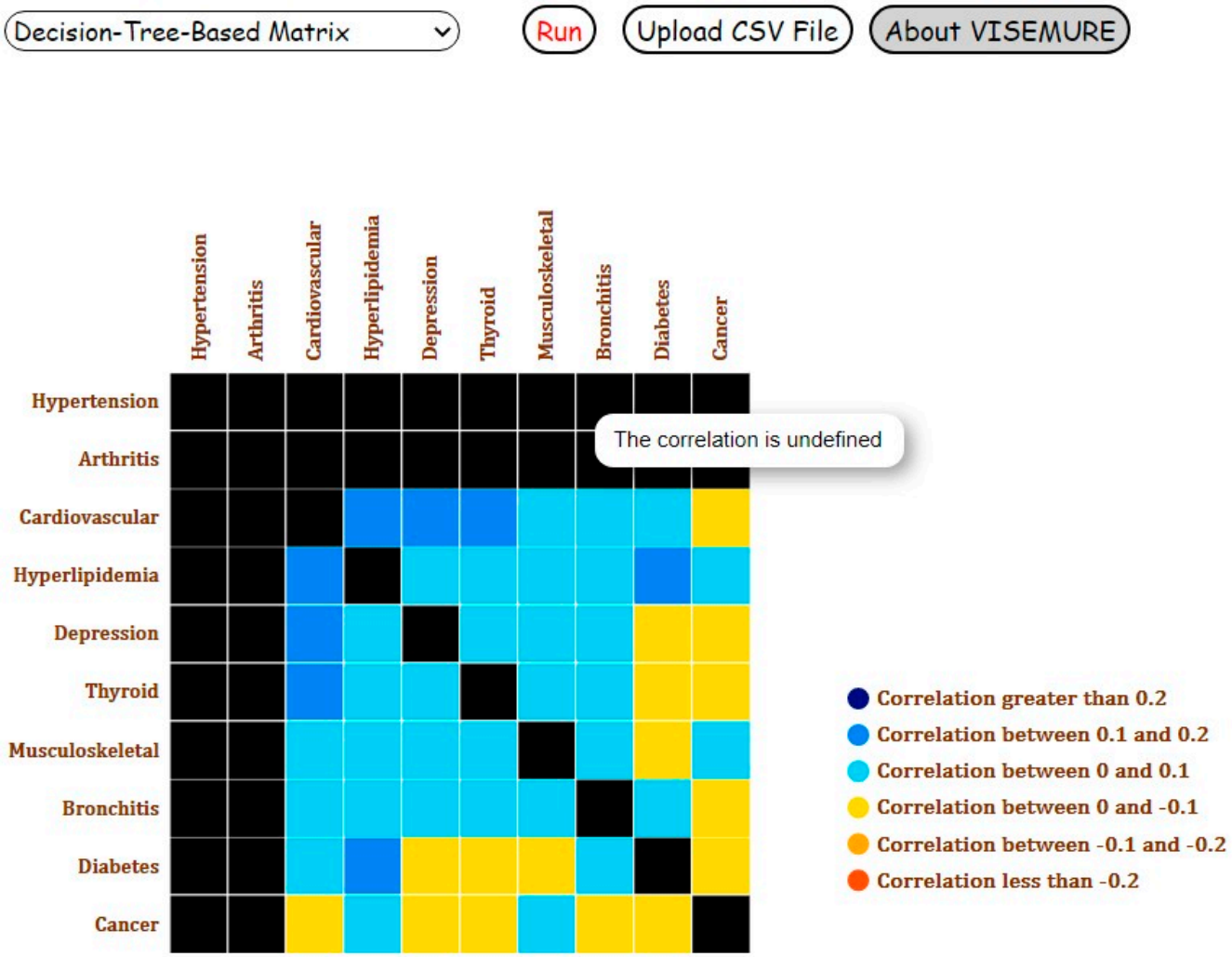

Figure 8. Screenshot of the Decision-Tree-Based Correlation Matrix for Analysis 3 with 'Elder' and 'Female' categories and the presence of hypertension and arthritis selected.

When we compare the two classifiers with count-based conditional probability in Table 4, the estimated probabilities are close to each other (e.g., diabetes, emphasized in the table), suggesting accuracy in our outputs. We use all of these models to stratify different covariates in the data. In other words, although the models improve prediction of the prevalence of and correlations between chronic diseases, they are being used to investigate the relationships between the diseases and sociodemographic characteristics.

Table 4. A comparison between three algorithms: Conditional Probability (Count-Based Bar Chart), Decision Tree, and Binary Logistic Regression, by assessing the prevalence estimates based on the selections in Analysis 3.

	HT	DB	BC	HL	CC	CD	DP	AT	TD	MP
Conditional Probability (Count-Based)	1	**0.235**	0.171	0.468	0.165	0.331	0.085	1	0.233	0.220
Decision Tree	1	**0.221**	0.170	0.382	0.136	0.261	0.091	1	0.184	0.177
Logistic Regression	1	**0.222**	0.145	0.401	0.127	0.326	0.068	1	0.170	0.140

Abbreviations: HT = Hypertension, DB = Diabetes, BC = Bronchitis, HL = Hyperlipidemia, CC = Cancer, CD = Cardiovascular Disease, DP = Depression, AT = Arthritis, TD = Thyroid Disease, MP = Musculoskeletal Problem.

Furthermore, the two machine learning algorithms, namely, softmax regression and decision tree, predict the same correlation coefficients in the cases the user only makes one selection (Analysis 1). We examined the correlations between cardiovascular disease and other chronic diseases to compare the performance of the two models through the three analyses. As shown in Table 5, for Analysis 1 in which only one variable is selected, the correlations estimated by softmax regression and decision tree are the same. As the number of selections increases, the results obtained from the two models differ from each other for some pairs of the diseases because they smooth the estimates in different ways (logistic regression assumes additive effects where decision trees do not). Which assumptions the user finds preferable will depend on the task and on that user's prior knowledge. For instance, in Analysis 3, the correlation between cardiovascular disease and bronchitis estimated by the softmax regression differs from the correlation between these two diseases predicted by the decision tree. This difference may come from softmax regression not being a count-based model. It borrows information from the other examples that are not selected, especially in cases like Analysis 3 with multiple selections where the number of selected examples is low, and the model needs additional information. Since the goal of our VA system is to explore the associations between variables rather than improving the predictions or determining the best classifier, these slight differences are not problematic.

Table 5. A comparison between two machine learning models, Softmax Regression and Decision Tree, which are used for correlation estimation, for all three analyses. Cardiovascular disease is chosen as an example to compare the estimated correlations between this disease and the other nine diseases in the data.

	Type	HT	DB	BC	HL	CC	DP	AT	TD	MP
Analysis 1	SR Model	0.122	0.004	**0.087**	0.140	0.035	0.010	0.044	0.002	0.017
	DT Model	0.122	0.004	**0.087**	0.140	0.035	0.010	0.044	0.002	0.017
Analysis 2	SR Model	0.106	0.037	**0.06**	0.147	−0.029	0.124	0.064	0.051	0.037
	DT Model	0.108	0.038	**0.05**	0. 144	−0.029	0. 116	0.056	0.06	0.054
Analysis 3	SR Model	-	0.061	**0.206**	0.248	0.029	0.251	-	0.239	0.196
	DT Model	-	0.018	**0.076**	0.181	−0.035	0.140	-	0.127	0.056

Abbreviations: HT = Hypertension, DB = Diabetes, BC = Bronchitis, HL = Hyperlipidemia, CC = Cancer, CD = Cardiovascular Disease, DP = Depression, AT = Arthritis, TD = Thyroid Disease, MP = Musculoskeletal Problem, SR = Softmax Regression, DT = Decision Tree.

5. Discussion

In this paper, we have shown how visual analytics systems can be used to explore patterns of multimorbidity. To achieve this, we have described the development process of VISEMURE, a VA system designed to satisfy the requirements of healthcare researchers in making sense of multimorbidity. VISEMURE incorporates a wide range of statistical and machine learning techniques and integrates them seamlessly with interactive data visualizations.

Using the DELPHI data, we have demonstrated how a health researcher can use VISEMURE to better understand the relationships among patient characteristics, existing disease states, and patterns of multimorbidity. The system is thus able to answer many questions that health researchers may have about these relationships in a way that affords them a great deal of freedom in terms of what characteristics to consider or exclude and on what diseases to focus as outcomes.

There is little research focusing on elaborate and interactive visualizations for enhancing the exploration of multimorbidity patterns [34]. Investigations in this area are mostly represented through static charts and tables that do not enable users to filter, select, control, and customize data points [1,35,36]. There are some interactive visualization systems that provide valuable healthcare insights by investigating the effects of patient characteristics and risk factors on the prevalence, incidence, or mortality of diseases [37]. However, such tools analyze only *one health outcome at a time*. Previous work on multimorbidity often

simplifies the outcome to a count of the number of chronic diseases present without distinguishing between them, although some work has investigated particularly common co-occurrence patterns [30]. This is the key difference between the VA system designed in this paper and other applications: VISEMURE allows users to investigate the distribution of a *multivariate* outcome, i.e., the joint occurrence of chronic diseases, in a way that supports a much richer set of questions about how those diseases are related to each other. This opens the path for a more detailed understanding of why chronic diseases co-occur, which in turn may lead to improved prevention and treatment strategies.

6. Limitations and Future Directions

The main limitation of this work stems from a design choice that we made regarding the type of analysis that VISEMURE is intended to support. Our system supports exploratory analysis, sometimes referred to as "descriptive epidemiology", rather than confirming facts that generalize beyond the data that are being analyzed. Exploratory analyses are crucial for understanding what is happening within a population of interest, and they are crucial for developing hypotheses around which relationships among risk factors, pre-existing conditions, and multimorbidity patterns may generalize to other settings. The proof-of-concept case study that we presented using the DELPHI data is designed to describe the DELPHI population and demonstrate the utility of the VISEMURE approach rather than to create broadly generalizable knowledge about multimorbidity.

To move beyond exploratory/descriptive tasks requires careful attention to issues of bias (e.g., what populations are or are not represented in the data) and variance (e.g., assessing confidence and statistical significance of findings). Issues of bias in data are well-known in epidemiology, and health researchers are trained to mitigate the bias present in data through prior knowledge and modelling. One future direction for VISEMURE would be to allow end-users to more finely adjust what information is used to control bias, perhaps by offering more flexible modelling options. The issue of variance or confidence is tied to the idea of statistical significance; the best way to address statistical significance in an interactive setting has been explored but is still an open area of research [38,39]. Developing a methodology for interactively mitigating bias and understanding variance so that VISEMURE can be used for a wider variety of tasks will be a focus of our future work.

Another avenue for future work would be to provide a richer view of the joint distribution of outcomes (conditional on patient characteristics and pre-existing conditions). We have used correlation to describe pairwise relationships among the different diseases, but it is possible that three-way or higher-order relationships are important in understanding the distribution of patterns of multimorbidity. Making sense of these more complex relationships would require substantially more development of the visual analytics tool in order to help the user to understand their salience and, thereby, to make them useful for sensemaking.

7. Conclusions

Multimorbidity is a growing healthcare challenge, especially for older adults, and results in greater vulnerability, higher risk of functional decline and disability, and higher mortality. Focusing on chronic diseases individually no longer meets the needs of patients or healthcare providers in preventing and managing these chronic conditions. A holistic approach to chronic diseases and their associations with sociodemographic characteristics and risk factors is needed to design effective prevention and control strategies. Therefore, we created a system for analyzing and exploring multimorbidity prevalence and associations in a visual, interactive manner. Unlike many studies in the area of multimorbidity whose results are shown through simple charts, tables, and flowcharts, our VA system allows users to interact with dynamic subsets of data and select a set of chronic diseases, and specific categories of age, gender, and socioeconomic scores for investigation.

The data visualizations in our system can be repurposed for other tasks in the area of healthcare or other disciplines where high-dimensional joint distributions of random

variables are important to understand. The system can also apply other statistical and machine learning models for prevalence and correlation estimation, and it can interpret more data with more available features.

VISEMURE is novel in the way it includes several statistical and machine learning techniques and integrates data analysis with interactive visualization to facilitate making sense of EMR data collected from patients with multimorbid diseases, which has never been attempted before. The design process established in this research will lead to the emergence of best practices for designing similar systems.

Author Contributions: Conceptualization, M.S.N., D.J.L., K.S., S.S.A.; methodology, M.S.N., D.J.L., K.S.; Software, M.S.N.; Validation, D.J.L., M.S.N., K.S., S.S.A.; Formal Analysis, M.S.N., D.J.L.; Data Curation, D.J.L., M.S.N.; writing—original draft preparation, M.S.N.; writing—review and editing, M.S.N., D.J.L., K.S., S.S.A.; Visualization, M.S.N., D.J.L., K.S.; supervision, D.J.L., K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partially funded by the Natural Sciences and Engineering Research Council of Canada.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Research Ethics Board of The University of Western Ontario (protocol code 114782, 2 October 2019).

Informed Consent Statement: This is a secondary data analysis of de-identified data. Patient consent was waived due to infeasibility of obtaining consent, low probability of re-identification, and low risk of harm to participations should re-identification occur.

Data Availability Statement: The data are held by the DELPHI team and are not publicly available. https://www.schulich.uwo.ca/familymedicine/research/csfm/research/current_projects/delphi.html.

Acknowledgments: We would like to thank the DELPHI team for their support for this project.

Conflicts of Interest: The authors declare that there is no conflict of interest.

References

1. Schäfer, I.; Kaduszkiewicz, H.; Wagner, H.-O.; Schön, G.; Scherer, M.; van den Bussche, H. Reducing Complexity: A Visualisation of Multimorbidity by Combining Disease Clusters and Triads. *BMC Public Health* **2014**, *14*, 1285. [CrossRef] [PubMed]
2. Fortin, M.; Bravo, G.; Hudon, C.; Vanasse, A.; Lapointe, L. Prevalence of Multimorbidity among Adults Seen in Family Practice. *Ann. Fam. Med.* **2005**, *3*, 223–228. [CrossRef] [PubMed]
3. Navickas, R.; Petric, V.-K.; Feigl, A.B.; Seychell, M. Multimorbidity: What Do We Know? What Should We Do? *J. Comorb.* **2016**, *6*, 4–11. [CrossRef] [PubMed]
4. Bähler, C.; Huber, C.A.; Brüngger, B.; Reich, O. Multimorbidity, Health Care Utilization and Costs in an Elderly Community-Dwelling Population: A Claims Data Based Observational Study. *BMC Health Serv. Res.* **2015**, *15*, 23. [CrossRef] [PubMed]
5. Wallace, E.; Salisbury, C.; Guthrie, B.; Lewis, C.; Fahey, T.; Smith, S.M. Managing Patients with Multimorbidity in Primary Care. *BMJ* **2015**, *350*, h176. [CrossRef]
6. Søndergaard, E.; Willadsen, T.G.; Guassora, A.D.; Vestergaard, M.; Tomasdottir, M.O.; Borgquist, L.; Holmberg-Marttila, D.; Olivarius, N.d.F.; Reventlow, S. Problems and Challenges in Relation to the Treatment of Patients with Multimorbidity: General Practitioners' Views and Attitudes. *Scand. J. Prim. Health Care* **2015**, *33*, 121–126. [CrossRef]
7. WHO. Projections of Mortality and Causes of Death, 2016 to 2060. Available online: http://www.who.int/healthinfo/global_burden_disease/projections/en/ (accessed on 18 February 2021).
8. Lang, T.; Shadmi, E.; Agmon, M. Electronic Health Records Use in Primary Care of Patients with Multimorbidity. *Int. J. Integr. Care* **2019**, *19*, 589. [CrossRef]
9. Rostamzadeh, N.; Abdullah, S.S.; Sedig, K. Data-Driven Activities Involving Electronic Health Records: An Activity and Task Analysis Framework for Interactive Visualization Tools. *Multimodal Technol. Interact.* **2020**, *4*, 7. [CrossRef]
10. Delamarre, D.; Bouzille, G.; Dalleau, K.; Courtel, D.; Cuggia, M. Semantic Integration of Medication Data into the EHOP Clinical Data Warehouse. *Stud. Health Technol. Inform.* **2015**, *210*, 702–706. [PubMed]
11. Abramson, E.L.; Barrón, Y.; Quaresimo, J.; Kaushal, R. Electronic Prescribing within an Electronic Health Record Reduces Ambulatory Prescribing Errors. *Jt. Comm. J. Qual. Patient Saf.* **2011**, *37*, 470–478. [CrossRef]
12. Abdullah, S.S.; Rostamzadeh, N.; Sedig, K.; Garg, A.X.; McArthur, E. Visual Analytics for Dimension Reduction and Cluster Analysis of High Dimensional Electronic Health Records. *Informatics* **2020**, *7*, 17. [CrossRef]

13. Abdullah, S.S.; Rostamzadeh, N.; Sedig, K.; Garg, A.X.; McArthur, E. Multiple Regression Analysis and Frequent Itemset Mining of Electronic Medical Records: A Visual Analytics Approach Using VISA_M3R3. *Data* **2020**, *5*, 33. [CrossRef]
14. Tang, P.C.; McDonald, C.J. Electronic health record systems. In *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*; Shortliffe, E.H., Cimino, J.J., Eds.; Health Informatics; Springer: New York, NY, USA, 2006; pp. 447–475, ISBN 978-0-387-36278-6.
15. Christensen, T.; Grimsmo, A. Instant Availability of Patient Records, but Diminished Availability of Patient Information: A Multi-Method Study of GP's Use of Electronic Patient Records. *BMC Med. Inform. Decis. Mak.* **2008**, *8*, 12. [CrossRef]
16. Nicholson, K.; Terry, A.L.; Fortin, M.; Williamson, T.; Bauer, M.; Thind, A. Examining the Prevalence and Patterns of Multimorbidity in Canadian Primary Healthcare: A Methodologic Protocol Using a National Electronic Medical Record Database. *J. Multimorb. Comorb.* **2015**, *5*, 150–161. [CrossRef] [PubMed]
17. Zheng, H.; Ryzhov, I.O.; Xie, W.; Zhong, J. Personalized Multimorbidity Management for Patients with Type 2 Diabetes Using Reinforcement Learning of Electronic Health Records. *Drugs* **2021**, *81*, 471–482. [CrossRef] [PubMed]
18. Melchiorre, M.G.; Lamura, G.; Barbabella, F. EHealth for People with Multimorbidity: Results from the ICARE4EU Project and Insights from the "10 e's" by Gunther Eysenbach. *PLoS ONE* **2018**, *13*, e0207292. [CrossRef] [PubMed]
19. Rind, A.; Wagner, M.; Aigner, W. Towards a Structural Framework for Explicit Domain Knowledge in Visual Analytics. In Proceedings of the 2019 IEEE Workshop on Visual Analytics in Healthcare (VAHC), Vancouver, BC, Canada, 20 October 2019; pp. 33–40. [CrossRef]
20. Marlin, B.M.; Kale, D.C.; Khemani, R.G.; Wetzel, R.C. Unsupervised Pattern Discovery in Electronic Health Care Data Using Probabilistic Clustering Models. In Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, Miami, FL, USA, 28–30 January 2012; ACM Press: Miami, FL, USA, 2012; p. 389.
21. Wetzel, R.C. The Virtual Pediatric Intensive Care Unit: Practice in the New Millennium. *Pediatr. Clin.* **2001**, *48*, 795–814.
22. Koh, H.C.; Tan, G. Data Mining Applications in Healthcare. *J. Healthc. Inf. Manag.* **2005**, *19*, 64–72. [CrossRef]
23. Simpao, A.F.; Ahumada, L.M.; Desai, B.R.; Bonafide, C.P.; Galvez, J.A.; Rehman, M.A.; Jawad, A.F.; Palma, K.L.; Shelov, E.D. Optimization of Drug-Drug Interaction Alert Rules in a Pediatric Hospital's Electronic Health Record System Using a Visual Analytics Dashboard. *J. Am. Med. Inform. Assoc.* **2014**, *22*, 361–369. [CrossRef]
24. Saffer, J.D.; Burnett, V.L.; Chen, G.; van der Spek, P. Visual Analytics in the Pharmaceutical Industry. *IEEE Comput. Graph. Appl.* **2004**, *24*, 10–15. [CrossRef]
25. Parsons, P.; Sedig, K.; Mercer, R.E.; Khordad, M.; Knoll, J.; Rogan, P. Visual Analytics for Supporting Evidence-Based Interpretation of Molecular Cytogenomic Findings. In Proceedings of the 2015 Workshop on Visual Analytics in Healthcare, Chicago, IL, USA, 25 October 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 1–8.
26. Abdullah Sheikh, S. Visual Analytics of Electronic Health Records with a Focus on Acute Kidney Injury. Ph.D. Thesis, The University of Western Ontario, London, ON, Canada, 2020. Available online: https://ir.lib.uwo.ca/etd/7086 (accessed on 20 March 2021).
27. Williamson, T.; Green, M.E.; Birtwhistle, R.; Khan, S.; Garies, S.; Wong, S.T.; Natarajan, N.; Manca, D.; Drummond, N. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann. Family Med.* **2014**, *12*, 367–372. [CrossRef]
28. Jeblee, S.; Khan Khattak, F.; Crampton, N.; Mamdani, M.; Rudzicz, F. Extracting relevant information from physician-patient dialogues for automated clinical note taking. In Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019), Hong Kong, China, 3 November 2019.
29. Birtwhistle, R.V. Canadian Primary Care Sentinel Surveillance Network. *Can. Fam Physician* **2011**, *57*, 1219–1220. [PubMed]
30. Nicholson, K. Multimorbidity among Adult Primary Health Care Patients in Canada: Examining Multiple Chronic Diseases Using an Electronic Medical Record Database. Ph.D. Thesis, The University of Western Ontario, London, ON, Canada, 2017. Available online: https://ir.lib.uwo.ca/etd/4483 (accessed on 2 January 2021).
31. Analysis. Available online: https://www150.statcan.gc.ca/n1/pub/75-202-x/2010000/analysis-analyses-eng.htm (accessed on 8 March 2021).
32. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, TX, USA, 28 June–3 July 2010.
33. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
34. PheWAS-ME: A Web-App for Interactive Exploration of Multimorbidity Patterns in PheWAS—PubMed. Available online: https://pubmed.ncbi.nlm.nih.gov/33051675/ (accessed on 8 March 2021).
35. Leng Low, L.; Heng Kwan, Y.; Shi Min Ko, M.; Teng Yeam, C.; Shu Yi Lee, V.; Boon Tan, E.; Thumboo, J. Epidemiologic Characteristics of Multimorbidity and Sociodemographic Factors Associated with Multimorbidity in a Rapidly Aging Asian Country. *JAMA Netw. Open* **2019**, *2*, e1915245. [CrossRef]
36. Raghupathi, W.; Raghupathi, V. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach to Public Health. *Int. J. Environ. Res. Public Health* **2018**, *15*, 431. [CrossRef]
37. Data Visualizations. Available online: http://www.healthdata.org/results/data-visualizations (accessed on 8 March 2021).

38. Xiaoying, P.; Matthew, K. ; Matthew, K. The Garden of Forking Paths in Visualization: A Design Space for Reliable Exploratory Visual Analytics: Position Paper. In Proceedings of the IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV), Berlin, Germany, 21 October 2018.
39. Foster, D.P.; Stine, R.A. α-investing: A procedure for sequential control of expected false discoveries. *J. R. Stat. Soc. Ser. B* **2008**, *70*, 429–444. [CrossRef]