Social Networks in Healthcare

Alind Jain, Namratha Manikonda, New York University

Social Networks between organisations and individuals can be made for information exchange, assistance and knowledge sharing. The method of connection is dependent upon their activity and are made into certain groups. a) Global social networks like FB, Twitter, etc b) Social networks pertaining to healthcare like healthcare forums, etc. c) Social networks for Scientific research or advancement like iMedExchange, are all just the tip of the iceberg. Social networks have their advantages and disadvantages which mainly pertain to information which can be either shared willingly or taken unprofessionals which can lead to breach in patient data, violation of personal professional boundaries, etc. Networks in Healthcare can also be created for communicable diseases during an Outbreak, which tracks patient #0 as well as how the disease is spread.

1 Introduction

Networks in Healthcare can be created to understand how the disease spreads as well as create relationships for eg. Social Networks to create the spread of Obesity [1] and the dynamics of smoking in large social networks [2]. We will look at Social Networks created between Healthcare providers like hospitals, nursing homes and Healthcare workers like doctors and nurses.
There has been a tremendous amount of work done to increase information gathering and sharing between healthcare departments and professionals but we still have not reached our potential because we are more scared of the disadvantages of data sharing than we are of its advantages. In this project, we will also focus on predicting cardiovascular heart diseases based on a certain set of variables and see how social networks can help in spreading awareness on prevention of cardiovascular heart diseases as well as other communicable diseases. Electronic health information exchange (HIE) allows sharing of Health records and confidential information with other healthcare providers.

2 Methodology

2.1 Data Sources

We take a look at data from the World Health Organisation which has both communicable disease data and cardiovascular data. The various communicable diseases that they record are Malaria, Tuberculosis, Hep-B, and HIV. Ourworldindata.org provides an annual number of deaths by cause for Europe and North America. Data.cms.gov provided Provider Information, Long Term Care

Hospitals and The Doctors and Clinicians national which helped us create Social Network healthcare graphs between providers and workers. We will look at a graph for the New York State healthcare facilities and speculate as to how these are connected. We also refer to the Framingham Heart Study Dataset [8] which is the study of instances and prevalence of CVDs, its risks and trends over time. It is a study conducted by the National Heart Institute (NIH). It has data of patients from Framingham,Massachusetts.

2.2 Analysis

Cardiovascular heart diseases are one of the most significant causes of death worldwide in this day and time. Predicting potential CHD in a person in the early stages could benefit the individual immensely. People at risk of CHDs can make lifestyle changes and take precautionary medicines to help counteract risks of having a heart disease. We will use the Framingham Heart Study dataset that has various risk factors that may lead to CHDs and perform logistic regression on this data to help predict the possibility of an individual getting a CHD in the near future. Various studies have shown that machine learning techniques can be used to predict and assist in making effective decisions in this field of study[10]. We need to select features that we know have a direct impact on heart health. Feature selection is probably one of the most important aspects of building an accurate model. Another important facet of this project would be the data used. We know data that isn't skewed as this would give us biassed predictions which would decrease the precision of our model.

Additionally, we take a look at the data visualisation for communicable diseases like Malaria, Tuberculosis, Hep-B and HIV data and notice the trend. We look at them individually and for Europe and North America we look at them as whole because they have the most advanced Healthcare Social Networks. We analyse the Social networks created by Healthcare workers like Doctors and nurses and then we speculate how the long term healthcare hospitals are connected with smaller healthcare providers like nursing homes.

HealthIT provides data for Hospital Public Health Reporting,

3 Results

We first preprocess the Framingham dataset [9] by removing any null values and check for outliers. After the data is preprocessed, we select the features that we will use to predict the dependent variable which is the potential risk of a person having a CHD in the next ten years. We select features based on the correlation matrix which we have depicted in the form of a heatmap as shown in Fig 1, and use the p-values and the correlation coefficient values.The features we selected to use for prediction are Gender, Age, Amount of cigarettes smoked per day, glucose level, their blood pressure and their cholesterol levels. We then build the logistic regression model using these features by splitting the dataset into a training set and test set in the ratio

80:20. We then check the model for its accuracy and check its confusion matrix which gives us the results of the model.

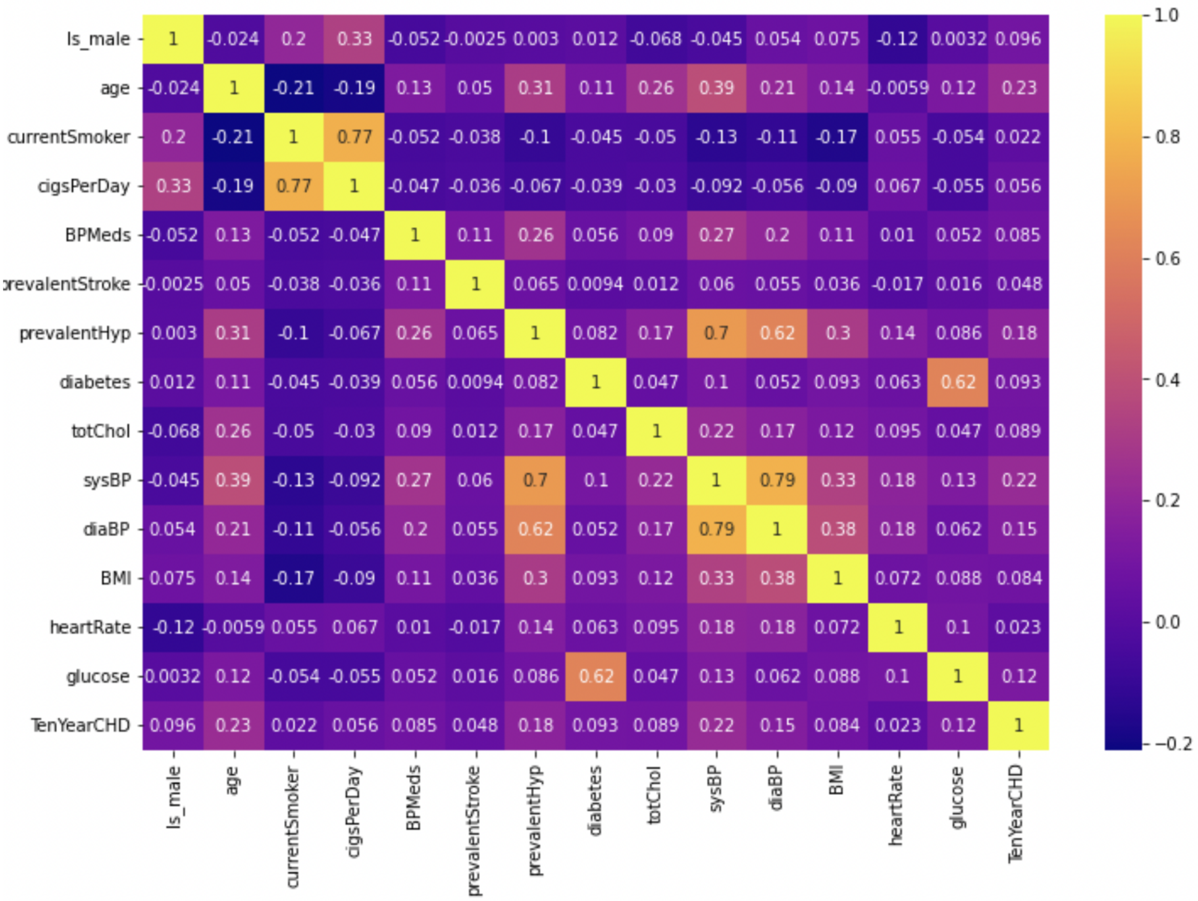


Fig 1. Feature Correlation Heatmap

The logistic regression model we built has an accuracy of 85.73%. We have the confusion matrix for the model below shown as Fig 2:

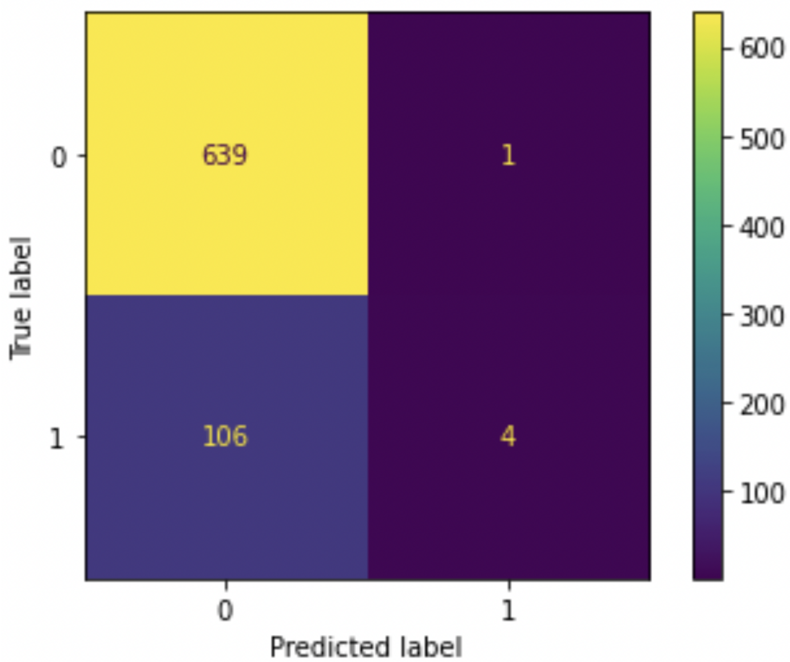


Fig 2. Logistic Regression Model Confusion Matrix

We can look at the graph in Fig 3 of Malaria deaths over the years and see that there is a gradual decline in the mortality over the years, one of the reasons being the communication between Social Networks created by Healthcare providers, where they share information about the curbing the disease and its treatment.

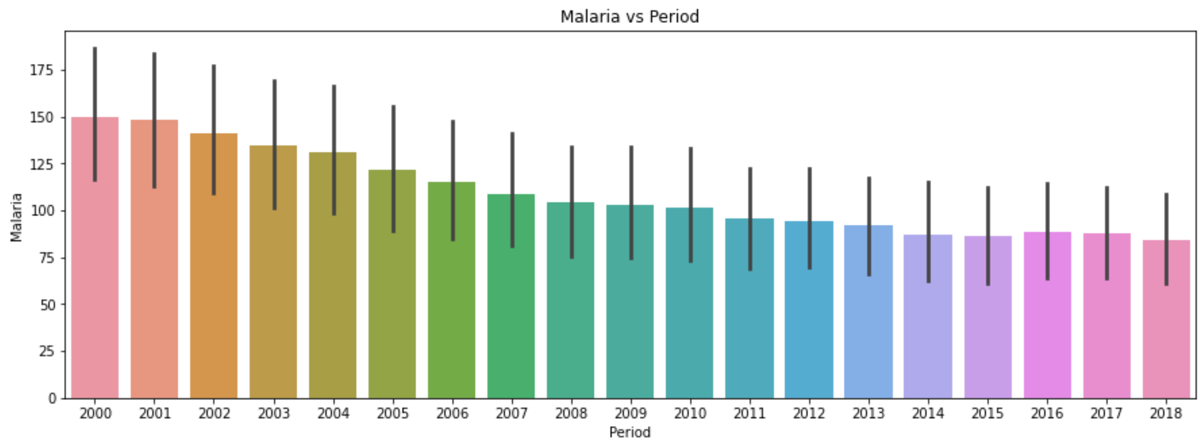


Fig 3. Malaria cases per 1000 uninfected people over the years

From figure 4 we can see the decline of deaths due to communicable diseases for both Europe and North America. Mortality rate due to communicable diseases for Europe is a lot higher than North America even when the Healthcare connectivity for Europe is far better than North America.

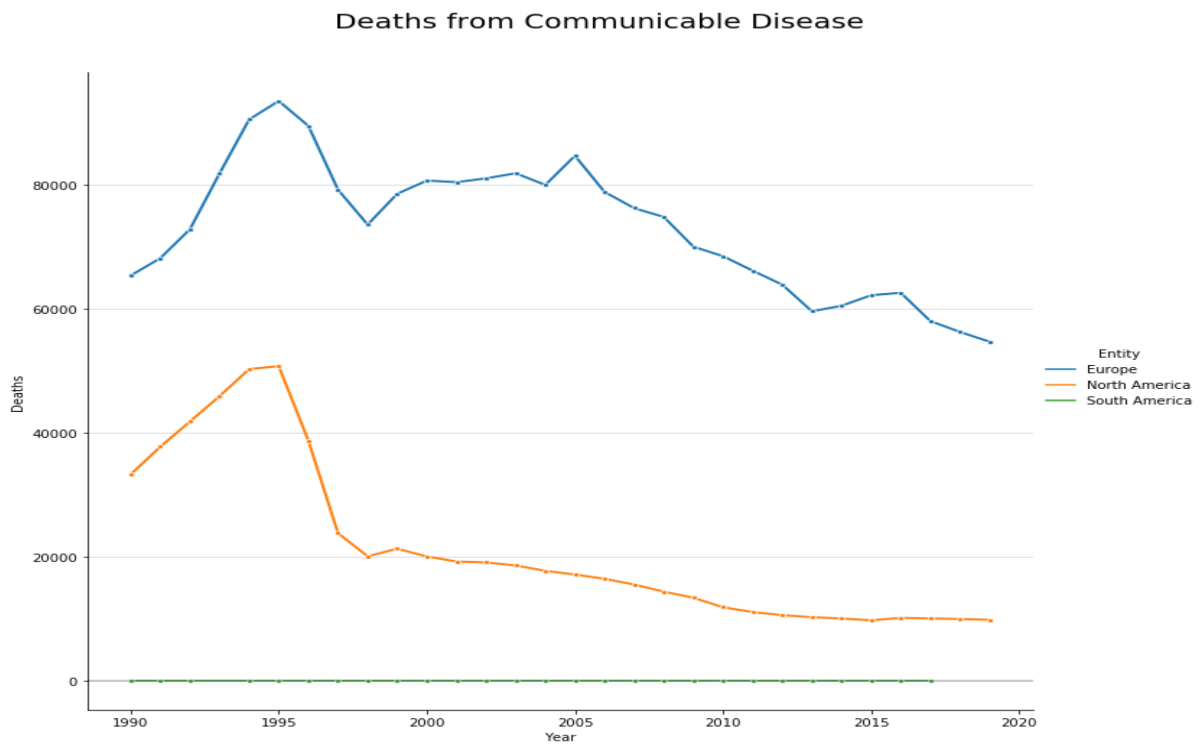


Fig 4. Deaths from Communicable Disease

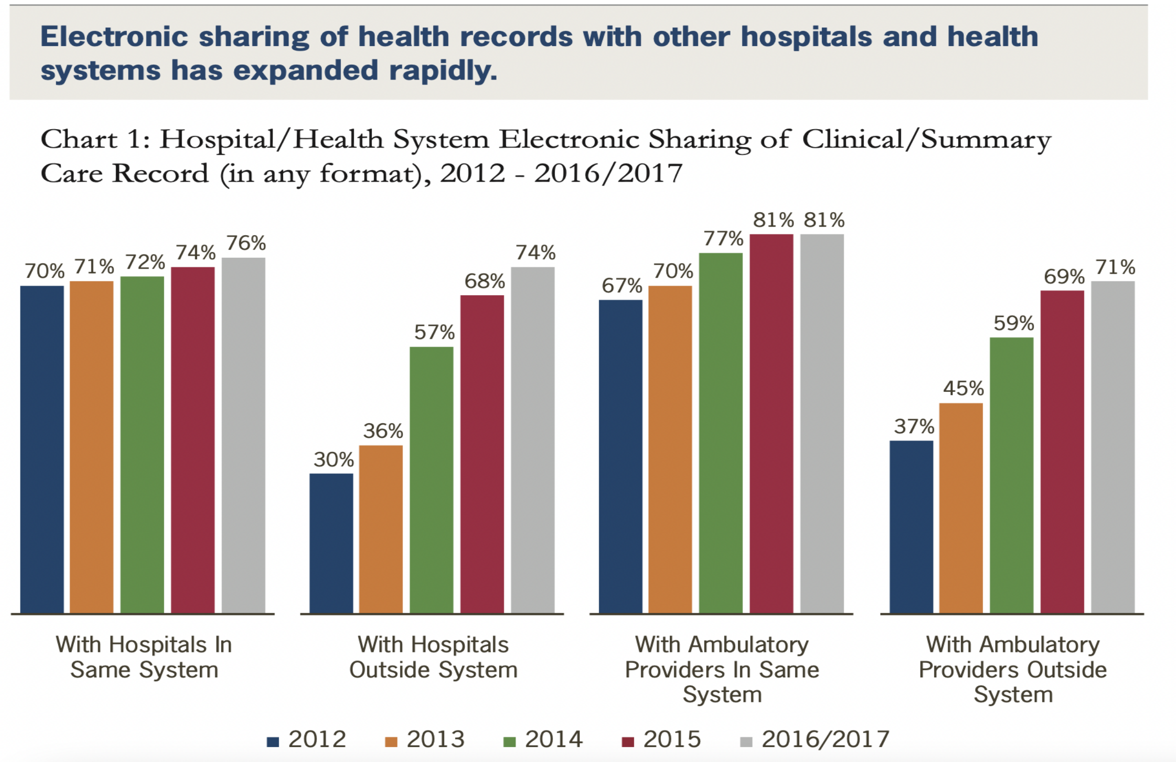


Fig 5. Electronic Sharing of Health Records [6]

From figure 4 and 5 we can say that the communicable deaths are going down and electronic sharing of Health data is going up. We cannot say that the decrease in mortality is due to exchange in health data because correlation is not causation from this data, but we can say that it is a factor.

Fig 6 shows a relationship between how the Tuberculosis outbreak happened with Patient 0 visiting a few places and the people getting infected with TB at those places who go to other places. We formed a Social Network of how TB spread in Houston.

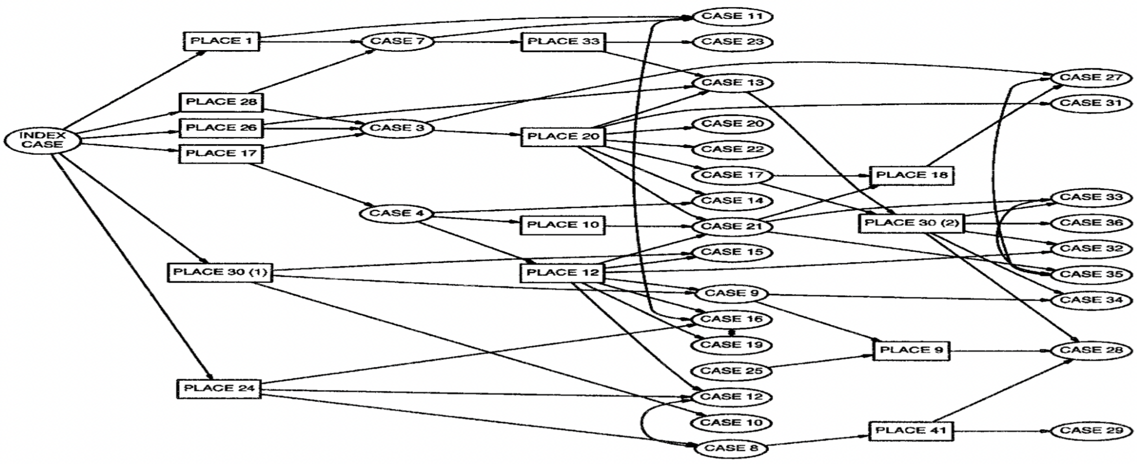


Fig 6. Relationships between Patients and Places during the Tuberculosis Outbreak in Houston [7].

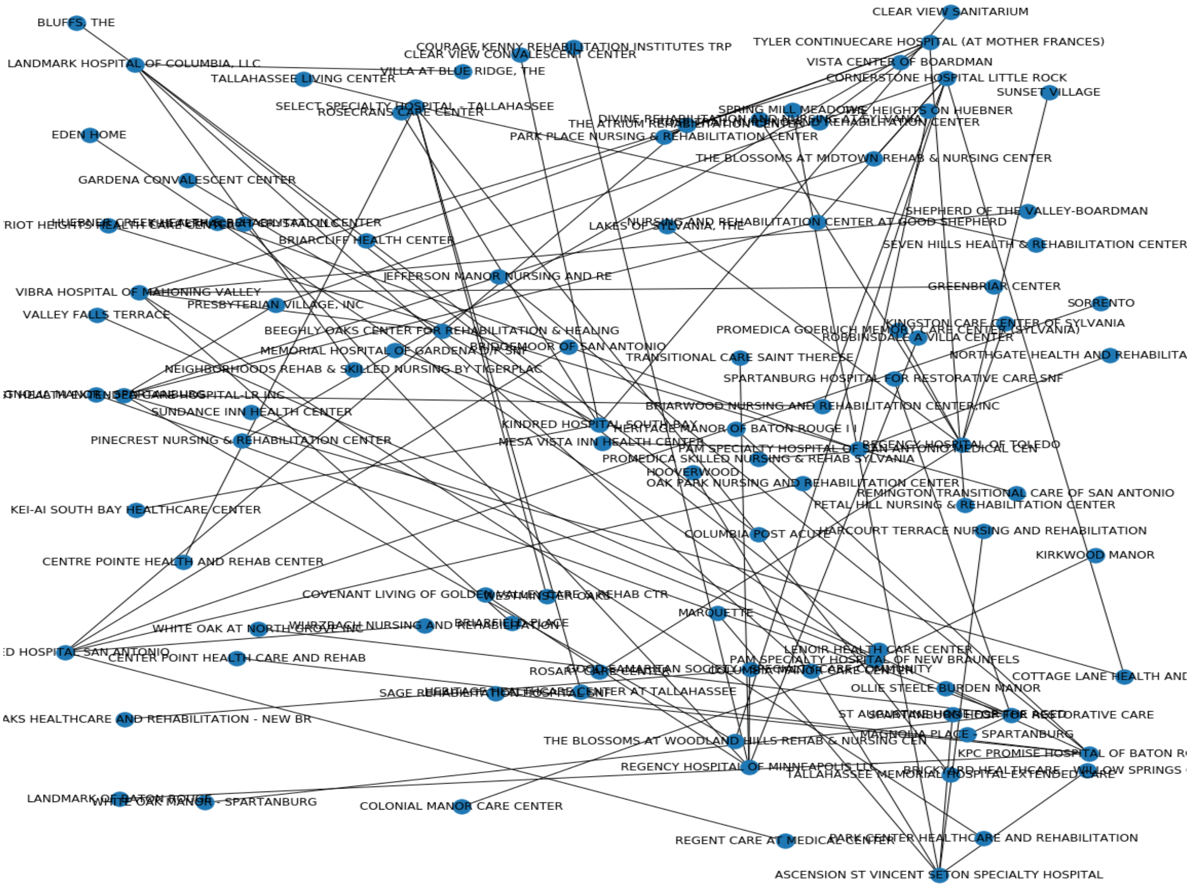


Fig 7. Social Network between Long Care Hospitals and Small Healthcare Providers nearby.

In Figure 7, we speculate how the Long Care Hospitals are connected to the Small Healthcare Providers which may have common healthcare workers and are in the vicinity of each other. We talk about how the nursing homes and small care facilities may be connected to two long care hospitals because of specialised consultations, emergency healthcare services as well as having similar providers

Figure 8 shows an existing Social Healthcare Network between Hospitals due to them having common Doctors and other healthcare workers who work or consult for two or more hospitals which forms a Healthcare Social Network.

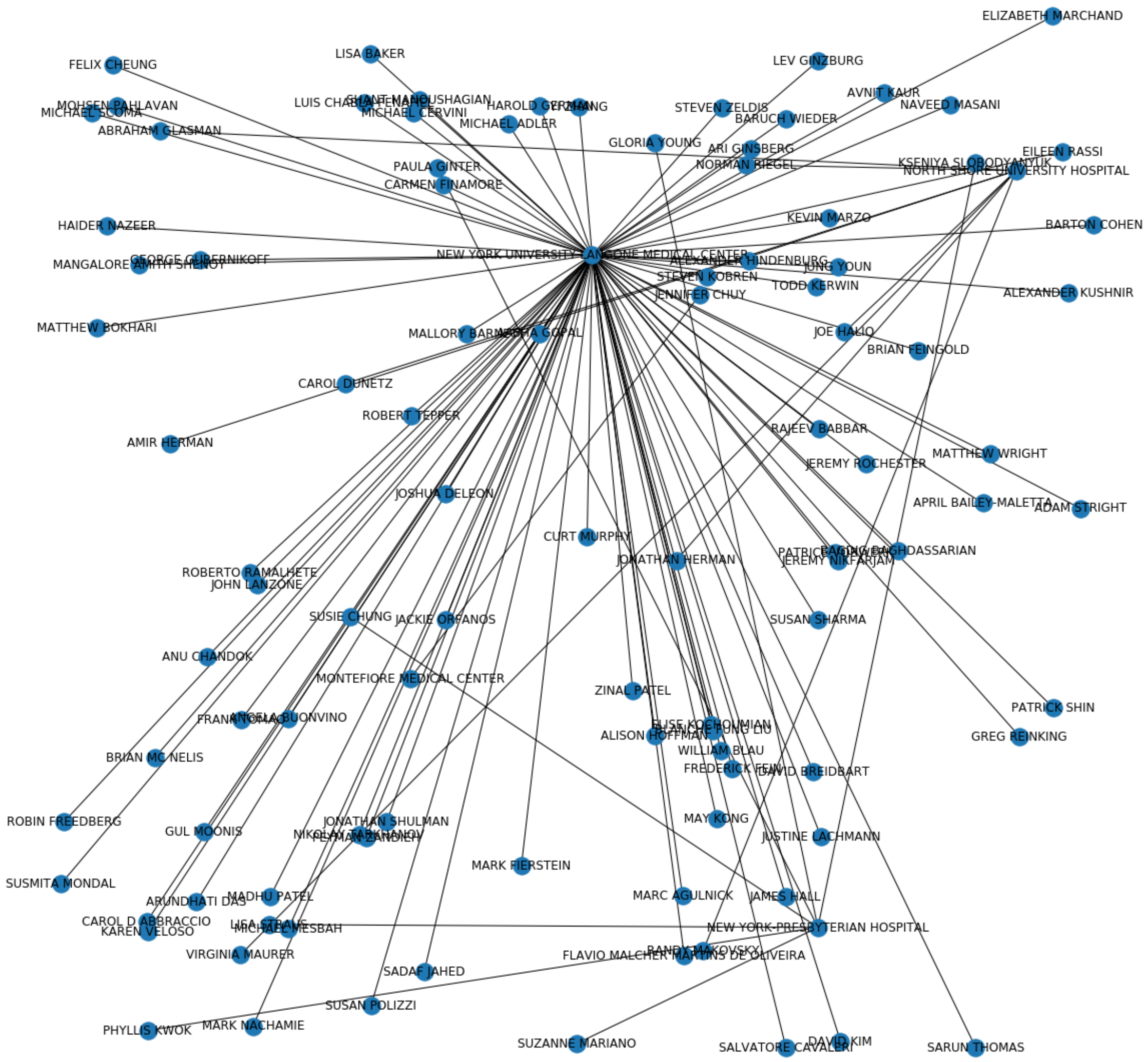
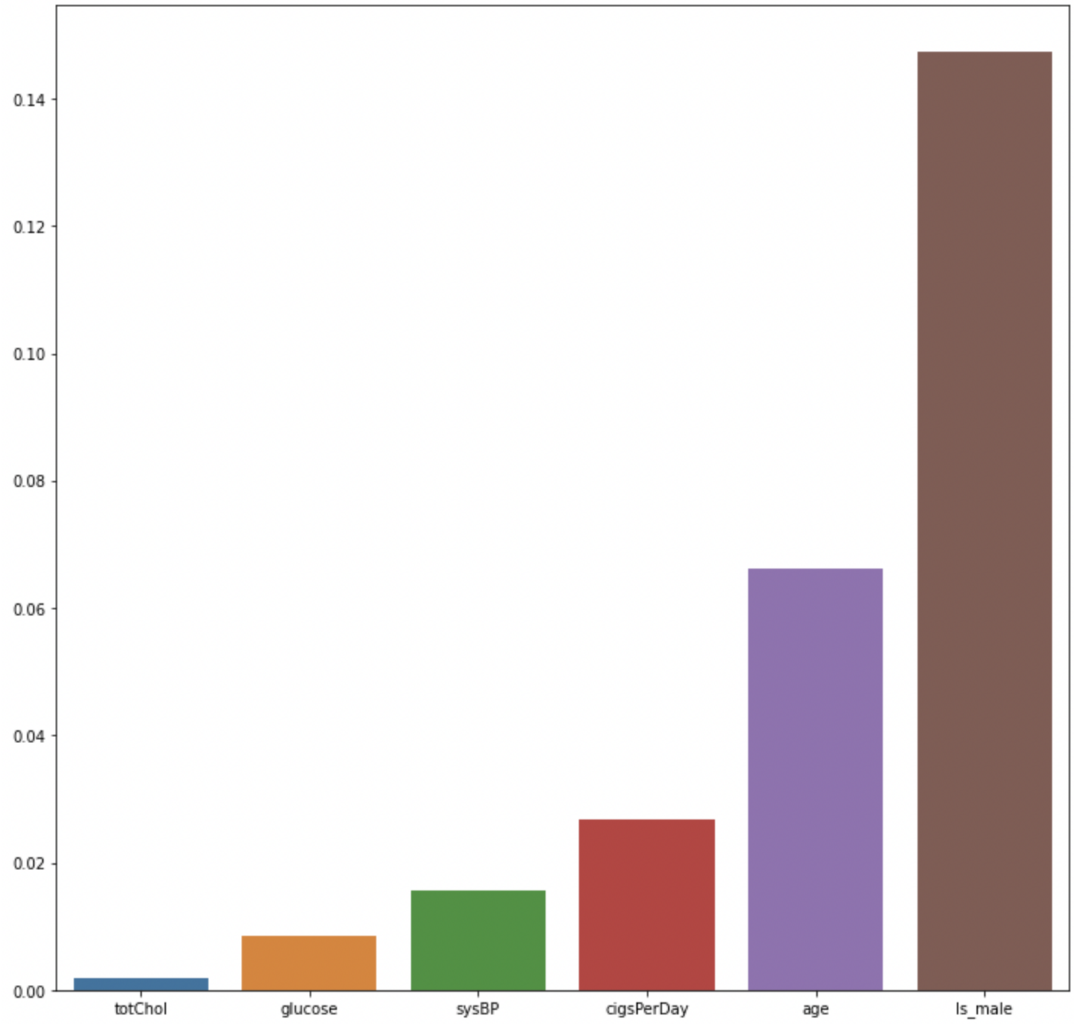


Fig 8. Social Network between Healthcare Workers and top 6 Healthcare Providers in New York State.

4 Discussion

As we can see from Fig 1, the features which play an active role in predicting whether or not an individual is at a potential risk of a CVD are Gender, Age, number of cigarettes smoked per day, glucose level, their blood pressure and their cholesterol levels. These values can be seen to have the maximum correlation in the heatmap shown in Fig 2. They also have p-values less than 0.05 which indicates that they are good predictors. After building the model with our preprocessed dataset, we observed an accuracy of around 85.73%. According to the confusion matrix, there were 639 true positives, 106 true negatives, 1 false positive and 4 false negatives. We also observe that in the features subset used to build the model, the features that had the maximum impact on our prediction were gender, age and number of

cigarettes smoked per day. We can see the coefficients graph for these features in Fig 9.
We also talk about the existing and theorised Social Networks which exist, the effects it may have on the decline of communicable diseases



5 Conclusion

According to the Odds Ratio we get from our Logistic Regression model, we can make the following conclusions: The odds of getting diagnosed with a CVD is 78.83% higher in males than females. Another conclusion is that with every one year increase in age, the odds of the individual contracting a CVD increases by 1.9%. We can also see that for every extra cigarette smoked by an individual the risk of getting a CVD goes up by 1.8%.
We see that there is a decline in the mortality rate for communicable diseases which may be in part to Electronic Sharing of Health Records which has increased over the years. But correlation does not equal causation so we need to perform more research before we can come to this conclusion. We have healthcare social networks which are built and show how the different healthcare providers be it Long CareHospitals, Nursing Homes, Small healthcare facilities, etc are connected with

each other. We see that the social network exists and can be improved upon to share electronic health records more efficiently.

6 References

[1] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007 Jul 26;357(4):370-9. doi: 10.1056/NEJMsa066082. Epub 2007 Jul 25. PMID: 17652652.
[2] Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med. 2008 May 22;358(21):2249-58. doi: 10.1056/NEJMsa0706154. PMID: 18499567; PMCID: PMC2822344.
[3] HealthIT.gov
[4] Ourworldindata.org
[5] data.cms.gov
[6] AMERICAN HOSPITAL ASSOCIATION ANNUAL SURVEY IT SUPPLEMENT BRIEF #2 MARCH 2018; https://www.aha.org/system/files/2018-03/sharing-health-information.pdf
[7] Klovdahl, A.S., Graviss, E.A. and Musser, J.M. (2002), "Infectious disease control: combining molecular biological and network methods", Levy, J.A. and Pescosolido, B.A. (Ed.) Social Networks and Health (Advances in Medical Sociology, Vol. 8), Emerald Group Publishing Limited, Bingley, pp. 73-99. https://doi-org.proxy.library.nyu.edu/10.1016/S1057-6290(02)80021-3
[8] https://biolincc.nhlbi.nih.gov/studies/framcohort/
[9] https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset/data
[10] https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8740989
[11] https://www.kaggle.com/datasets/utkarshxy/who-worldhealth-statistics-2020-complete
[12] B. Shneiderman, C. Plaisant and B. W. Hesse, "Improving Healthcare with Interactive Visualization," in *Computer*, vol. 46, no. 5, pp. 58-66, May 2013, doi: 10.1109/MC.2013.38.
[13] Thomas W. Valente. Copyright © 2010 by Oxford University Press, Inc. Social Networks and Health. Models, Methods, and Applications.
[14] Bernice A, Pescosolido and Judith A. Levy. The role of social networks in health, illness, disease and healing: the accepting present, the forgotten past, and the dangerous potential for a complacent future. In: Judith A, editor. Levy Social Networks and Health. Emerald Group Publishing Limited; 2002.
[15] Thompson LA, Black E, Duff WP, Paradise Black N, Saliba H, Dawson K. Protected health information on social networking sites: ethical and legal considerations. J Med Internet Res. 2011 Jan 19;13(1):e8.
[16] B.W. Hesse et al., “Social Participation in Health 2.0,” Computer, Nov. 2010, pp. 45-52.
[17] S. Afzal, R. Maciejewski, and D. Ebert, “Visual Analytics Decision Support Environment for Epidemic Modeling and Response Evaluation,” Proc. Conf. Visual Analytics Science and Technology (VAST 11), IEEE, 2011; http://ieeexplore.ieee. org/stamp/stamp.jsp?arnumber=06102457.
[18] José Alberto Benítez, José Emilio Labra, Enedina Quiroga, Vicente Martín, Isaías García, Pilar Marqués-Sánchez, Carmen Benavides, "A Web-Based Tool for Automatic Data Collection, Curation, and Visualization of Complex Healthcare Survey Studies including Social Network Analysis", Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 2579848, 8 pages, 2017.https://doi.org/10.1155/2017/2579848

[19] T. J. Carney and C. M. Shea, “Informatics metrics and measures for a smart public health systems approach: information science perspective,” Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 1452415, 12 pages, 2017.