1. SMART HOSPITAL REFERRAL SYSTEM

2. ABSTRACT

This study presents a method for constructing an expert system using a hospital referral problem as the primary focus. Choosing a medical care and aid institution that has a track record of providing quality care can make the difference between health improvement or health deterioration. The contrary difference has been contributed by several factors such as institutional characteristics, patient survival prospect, traveling distance, and complications. This has led patients to lose their lives on the way, especially pregnant mothers at the time of childbirth. Hospital referral standards come from the state of quality, intuition, and personal acquaintance. In institutional characteristics, the volume of patients or procedures is the most consistent predictor of in-hospital mortality and is broadly used as a hospital-selection criterion. Usually, large institutions have favorable characteristics, such as technical sophistication and more staffing, and they are usually preferred for a referral. Low-volume hospitals may be less inclined to turn down high-risk cases, large-volume hospitals attract more cases through physician or self-referral, and patients with opportunity and desire to be referred may be healthier because of several factors. This expert system will base on an algorithm that will predict and optimize decisions based on various factors from the patient and the hospitals to be referred to. Geographical characteristics play a major role in building a good hospital referral decision system, this is because often medical situations are urgent with time playing a critical role. A vivid example is patients situated in rural and remote areas, who to some prefer local hospitals despite being at high risk because proximity to medical aid is crucial for the survival of the patient. In a perfect scenario, each patient should be treated individually, with the decision process including not only their condition but also a tradeoff with the desired hospital features. In this case scenario, it is required to have an expert system with an algorithm that can help with complex decision making, especially when numerous factors are to be considered. The algorithm obtains knowledge by building a machine learning model classifier from a dataset that is a collection of labeled cases. As an output to a patient query, the algorithm gives a customized recommendation, using an optimization step to help the patient maximize the probability of the desired response. This desired response is the recommended hospital. With proper computation and formulation, this expert system can combine multiple factors to give hospital-selection decision support at an individual level. Therefore, this expert system will be able to solve the complex referral problem through knowledge in the form of the mathematical function from a classifier and apply optimization methods.

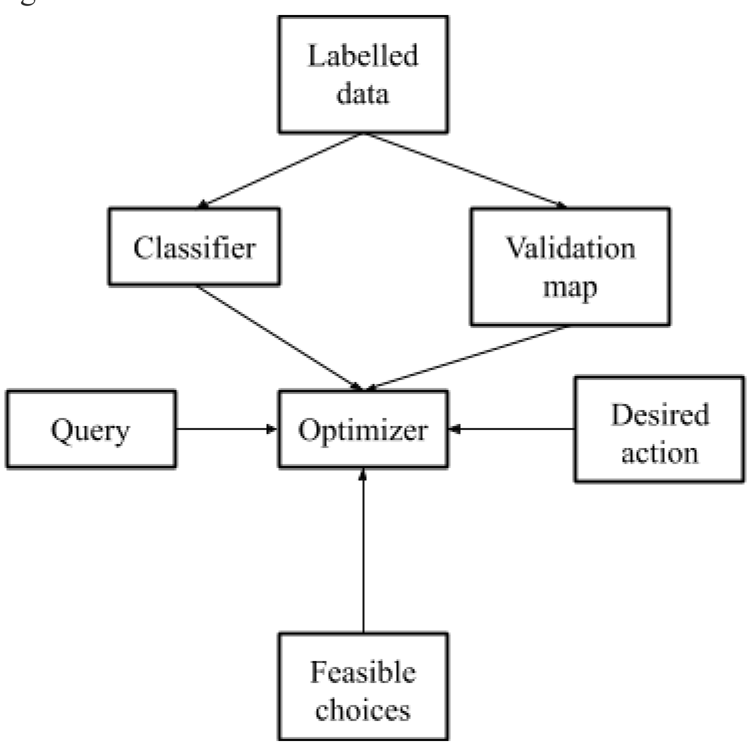
3. INTRODUCTION

The purpose of this expert system is to generate a suggested action that leads to a higher probability of saving lives by ensuring the patient is referred to the right institution. This type of problem can be found in many domains. Experts are often consulted because they know how to maximize the probability of the desired results while considering multiple and sometimes competing factors. The proposed algorithm can simulate these experts by recommending actions that maximize the probability of the desired result.

Many approaches can help to understand the constraints that lead to the desired outcome. The use of regression techniques can allow us to determine the relative importance of variables by observing the sign and magnitude of coefficients. Sensitivity analysis provides a way to observe how sensitive a result is to variations in the variables of interest, thus determining the importance of these variables.

These methods can identify important factors for achieving the desired outcome. However, they cannot tell us what to do, how to do it, and how to resolve trade-offs among alternatives. For example, the above methods can identify which interventions are most likely to get the patient to the right hospital, however, it usually takes an experienced expert to know how to choose among alternative interventions, and how to tailor the intervention to the needs of a specific patient.

The proposed algorithm is a decision tool that can provide suggestions by utilizing captured knowledge and optimizing the effectiveness of the chosen action. This algorithm can recommend an action decision based on multiple variables and the interactions among them. Example in the figure below



Instances with known outcomes are used to capture knowledge in the form of a predictive model (classifier) and a validation map that estimates the probability of the desired outcome for any patient/action pair. A query will activate an optimization method that finds the best course of action using the captured knowledge, feasible choices and information provided about the patient.

Communication between the decision support system and a user is required. The user provides information regarding a patient's characteristics and the maximal distance to a hospital that the patient can tolerate, and then the system can generate a customized hospital choice. This customized choice not only satisfies the given maximum tolerated distance but also identifies the hospital with the highest probability of the desired outcome. The maximum tolerated distance parameter should be decided by a patient and his/her doctor to ensure that the travel distance does not become a risk factor and is acceptable to the patient. For some healthier patients, this parameter value can be high. For an emergency case, this parameter must be very low.

Travel distance and survival probability are two important targets. The trade-off between different objectives can be addressed explicitly when the hospital choice decision is customized. In this study, the problem is first formulated as single-objective optimization and can be solved by an exhaustive search because of the small solution space (only the number of hospitals). In this optimization, a query provides a patient's characteristics to the system, such as age, admission type, comorbidities, and the maximum tolerated distance. The optimization process will combine the provided information with the captured knowledge to generate a customized hospital selection. The objective of the knowledge extraction tool is to find a hospital with the highest survival probability under the constraint of the maximum tolerated distance to a hospital. If we also want to consider other issues in the hospital-selection decision, such as the reduced likelihood of complications, then the problem is formulated as multi-objective optimization. We present the entire solution space to the user in an intuitive format so that the relative importance among the targets can be decided by the user. Due to the small solution space, presenting the predicted outcomes of each hospital in an organized way to a user is more effective and efficient than having the system decide on a single optimal choice.

A good decision-making system should also consider patient factors like tolerance. Empirical problems will arise if you only consider institutional factors and may get even tougher if other practical concerns like inpatient fees and the number of diagnostic cases.

To solve this problem, there is a need of having a Knowledge-Based System (KBS). This knowledge exists in the form of atomic facts about the domain of interest and rules for inferring new facts, but may also be in the form of graphs, trees, or networks. This data is stored in a specific location known as a knowledge base. Together with the inference engine, the knowledge base system is used to make inferences.

Another technique used in case-based reasoning, instead of using a knowledge base it uses case libraries which are all previously solved cases. The working mechanism for the solution for the new case involves finding the most similar problems in the case library.

Machine learning is applied in areas that require an expedition for knowledge. In turn, help to excavate knowledge from areas that seem cumbersome to comprehend. Machine learning has played a crucial role in various areas such as classification, diagnosis, and prognosis. Application of machine learning can be classified as supervised or unsupervised learning methods. Supervised learning methods learn a function that maps features that are the independent variable used to represent a case to the corresponding labels which can be dependent variables or typical outcomes. But due to cluster samples based on the similarity of variables, the labeled samples for unsupervised learning methods are not necessary.

Supervised learning methods that learn patterns in the form of rules provide a partially automatic method for knowledge accession in traditional expert systems example C4.5. Nevertheless, this requires an abundance of time and labor force to build and maintain the knowledge base. Another form of supervised learning involves mathematical functions which are induced patterns that can either be linear or nonlinear, in these functions knowledge is encoded. But on the contrary side, it is strenuous to use the knowledge in a mathematical function to assist in an action decision, such as a customized hospital-selection decision.

4. ANALYSIS/EVALUATE TECHNOLOGY

The algorithm relies on classifiers to capture knowledge whereby the prediction model is trained by the independent and dependent variables. This model is then expected to provide a score that is then converted to a probability, which is interpreted as the confidence level of the desired class prediction. Optimization is applied to maximize the confidence level of the desired class label. The following are the series of stages in the algorithm:

I. Design the dataset

The perspective of the solution to the problem depends on two types of independent variables, that is unchangeable and changeable variables. The unchangeable is mostly patient-based hence, given to us an example, diagnostic result, medical test result, and demographic data. The changeable variables, whose values can be changed. The recommendation can be made based on these variables

Variable	**Data type**
Patient age	Numeric
Patient sex	Male, Female
Patient admission type	Emergency, Urgent, Elective
Patient comorbidity severity	Numeric
Patient payment type	Insurance, Cash
Hospital ownership type	Government, Private
Hospital bed size	Numeric
Hospital surgery volume	Numeric
Hospital discharge volume	Numeric

The first set is constant and provided by the patient when querying. The optimal solution is the hospital with the most favorable descriptive variables that result in the highest optimum value that is the outcome with the highest probability.

Hospital data is collected manually from each individual then stored in the database to allow further usage in the future. The distance estimation from Google Maps could be used to compute road distance since it more accurately represents travel distance.

In this study, we require four datasets. The first dataset includes all patients whether surgery was performed or not. The second is designed for patients who do not have surgery. The third dataset includes patients who have received intensive care. The last dataset patients who have received intensive care due to hospital-acquired complications. Dependent variables, such as survival status and complication status, are usually used to find the predictors of hospital quality.

II. Build a predictive model.

The process of obtaining information in our model relies on supervised learning, which determines the relationship between independent and dependent variables. We used support vector machines (SVMs) as the classifier to construct a separating surface between point sets of different classes. There is an infinite number of surfaces that can perform the separation. To generalize to unseen points well, the SVM classifier finds the separating surface with the greatest margin, or the distance from a point to the surface, during the training process.

In our example, the surface separates negative from positive cases, with the negative area to the left of the plane ($d(x) < 0$), and the positive region to the right ($d(x) > 0$). If a test case falls to the right of the separation surface, the predicted result will be positive. In general, some points cannot be separated and are classified incorrectly. A high prediction score means a high probability that a patient will have the desired outcome. In other words, the confidence level of the desired outcome is high. On the other hand, if we can increase a data point's predictive score by changing certain independent variables (controllable), we may improve the probability to have the desired outcome.

Therefore, a predictive model using independent variables of patients' characteristics or uncontrollable variables, and their chosen hospitals' descriptive features or controllable variables, to predict whether or not a patient will survive (or be free from complication) during a hospital stay. Our goal is to increase this confidence level.

III. Extract recommended information

In this problem, we assume that the only way to improve a patient's expected outcome is to change hospitals. Optimization methods provide a scientific way to improve the confidence level to the desired outcome. The decision function can naturally become the source of the objective function since we want to maximize the confidence that the desired outcome occurs.

The idea of using optimization is intuitive. In real life, the probability of survival in different hospitals for the same patient will vary. After training, a classifier can estimate the score of survival or freedom-from-complication for a query patient in each hospital by an evaluation function. A good hospital will result in a higher predictive score, so the evaluation function value will be higher. In other words, the optimization method can help to identify such a hospital. The mathematical model can represent the hospital referral scenario in real life since a high-quality institution can prevent medical error and increase the chance of survival.

The survival function optimization is given as follows:

$$\underset{x2j}{\text{maximize}} \quad d(x1 \cup x2j)$$

$$\text{subject to} \quad \text{dist}(j,x) <= DL$$
$$x2j \in X2$$

where x1 is the characteristic variables of the query patient, and x2j is the set of descriptive variables describing the hospital j. dist(j, x) is the Euclidean distance from the patient to the hospital j. DL is the maximum tolerated traveling distance parameter given by the user'.

The purpose of the optimization process is to find the hospital with the most favorable descriptive features, such that the objective function is the maximum. Although a high-quality hospital can improve a patient's survival probability, the effect is sometimes limited. The physical condition of each patient is different and is a more important factor in deciding the survival probability. There are three possible situations when patients change their original chosen hospital to the referred one:

- Patients may move from predicted negative to less negative (less predicted probability of death).
- Patients may move from predicted negative to positive (predicted death to predicted survival).
- Patients may move from predicted positive to more positive (increasing the probability of the prediction of survival).

These methods can construct the character of a "perfect" hospital. However, in our application, this approach is impossible, since such a hospital does not necessarily exist. To maintain feasibility, the optimization should not create the value of the set of descriptive variables describing the hospital. Instead, we should evaluate the value of the objective function from all hospitals within the distance limit, and find one with the highest value of probability.

IV. Building a validation map

Support vector machines(SVM) are not probabilities. Despite being able to improve the predictive score by recommending a hospital to the patient's query. It is still the main priority to comprehend how much this has helped the patient. Using Platt's calibration method provides a computational solution, although we can observe the distribution of predictive scores corresponding to survival by learning the relationship between class labels and predictive scores.

The main purpose of the knowledge capturing (training) is to learn a decision function and probability transfer function from the data. In other words, knowledge is stored in these functions. In the single-objective optimization, given the query patient's characteristic variables and the maximum tolerated distance, the optimization will find the hospital with the highest survival probability satisfying the distance constraint. In multi-objective optimization, a patient does not need to give the maximum tolerated distance. Instead, the algorithm will give information including the survival probability, the freedom-from-complication probability, and the distance to each hospital. They comprise the patient-specific information and can be expressed visually, as a consumer report. If there are two identical patients, their hospital selection could be different because their ideas of the importance of each objective may vary.

V. Obtain results and objective optimization

In the following sections, we present the results using single- and multi-objective optimization. In the actual application of the single-objective optimization, the maximum tolerated distance should be decided by a user, and the returned optimal solution is customized. We varied this parameter to present results. In the application of multi-objective optimization, a user does not need to give a parameter. Instead, the user needs to choose the optimal solution in the solution space considering three desired targets. Similar to the single-objective optimization, we varied the distance target and discussed the user's decision considering the other two desired targets.

5. CONCLUSION

We proposed a new data mining process to construct an expert system, whereby data mining provides a framework for the extraction of important information from data. This has enabled utilization of the classifier to perform decision functions also improve reliability on an optimizer to choose the action that maximizes the confidence in the outcome. The structure of this process has provided many advantages including, automatic construction and maintenance which lead to low cost, it has improved flexibility in the estimation of real estimation, and lastly, the best solution can be found. The proposed process can be adapted to solve many types of problems.s.

6. REFERENCES

Knowledge Creation Opportunities in the Data Mining Process - IEEE Conference Publication, ieeexplore.ieee.org/abstract/document/1579639.

Guest Editorial Introduction to the Special Issue on Advances in Clinical and Health-Care Knowledge Management - IEEE Journals & Magazine, ieeexplore.ieee.org/abstract/document/1435412.

A Knowledge Creation Info-Structure to Acquire and Crystallize the Tacit Knowledge of Health-Care Experts - IEEE Journals & Magazine, ieeexplore.ieee.org/abstract/document/1435417.

Birkmeyer, John D., et al. "Hospital Volume and Surgical Mortality in the United States: NEJM." *New England Journal of Medicine*, 11 Apr. 2002, www.nejm.org/doi/full/10.1056/NEJMsa012337.

Birkmeyer, John D., et al. "Volume Standards for High-Risk Surgical Procedures: Potential Benefits of the Leapfrog Initiative." *Surgery*, Mosby, 25 May 2002, www.sciencedirect.com/science/article/abs/pii/S003960600153860X.

Chi, Chih-Lin, et al. "Building a Hospital Referral Expert System with a Prediction and

Optimization-Based Decision Support System Algorithm." *Journal of Biomedical Informatics*, Academic Press, 22 Oct. 2007, www.sciencedirect.com/science/article/pii/S1532046407001086.

"Data Mining Process." *Data Mining Process - an Overview | ScienceDirect Topics*, www.sciencedirect.com/topics/computer-science/data-mining-process.

Dr. Kupersmith is a Petersdorf Scholar-in-Residence. "Quality of Care in Teaching Hospitals: A Literature Review : Academic Medicine." *LWW*, journals.lww.com/academicmedicine/Fulltext/2005/05000/Quality_of_Care_in_Teaching_Hospitals__A.12.aspx?__cf_chl_jschl_tk__=e9145ee513b7801375365143db90fa1ea55ddfbc-1609687445-0-AaX90SUywUnXe_tp8DzNnGrIcHiND5t-G-LmSRLVE0Wd_5EkrG6_F9vdSNfY3QZ1cPLTZMVtdxzHhkXJOnmzzEo3cNsZFxlWABzk2ejtPQFUUKkBLhxhkHICmMrbJEUYsRRSXnC0usAxfUvdz2cKSuGjaOewidqSezUexVf9fJVdkwOudbArRgL3Ftd8Pl_aw-8joViJOMwbFwLvn-J7UXOfFBPNfpLyZYN8ZkhfvsELVjcFoMGxOz2Mv1ueycMCwW9b4_rY4TG-1vf3Tfo6ajUVW7hGOAOnF6Xl-Dd0VrxsXki_jtDDqMxnmlmDRrvnnioFTejjP-F2mhfVCIuCEdAl2yLzbiWawRgzRBZWMaXguEyOEsrm_BHnY7p3iBPsADX-Z_Jgl03clrPcGNF4BH_qpVpjN486t7XojHn11AESomF8XUGb1k3rbOiSnQs9DedYlWwDc2mGug9oZAmuLrHXBze1ft7U6pfkqTMkRNIc.

Halm, Ethan A., et al. "Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature." *Annals of Internal Medicine*, www.acpjournals.org/doi/abs/10.7326/0003-4819-137-6-200209170-00012.

Halm, Ethan A., et al. "Is Volume Related to Outcome in Health Care? A Systematic Review and Methodologic Critique of the Literature." *Annals of Internal Medicine*, www.acpjournals.org/doi/abs/10.7326/0003-4819-137-6-200209170-00012.

Jeroan J. Allison, MD. “Relationship of Hospital Teaching Status With Quality of Care and Mortality for Medicare Patients With Acute MI.” *JAMA*, JAMA Network, 13 Sept. 2000, jamanetwork.com/journals/jama/article-abstract/193061.

Shortliffe, Edward H., et al. “Computer-Based Consultations in Clinical Therapeutics: Explanation and Rule Acquisition Capabilities of the MYCIN System.” *Computers and Biomedical Research*, Academic Press, 20 Oct. 2003, www.sciencedirect.com/science/article/abs/pii/0010480975900099.