CSC 215-01 Artificial Intelligence (Fall 2020)

Final Project Proposal

Paper Title: One shot at AKI prediction

Project Type: Type A - Research Project

Team Members: Harshitha Onkar (Sac Id: 220262706), Gargi Prabhugaonkar (Sac Id: 301111525)

Problem Description/Abstract:

Acute kidney injury (AKI) is the most common postoperative complication among surgical patients. The incidence of postoperative AKI accounts for 18%–47% of total hospitalized AKI patients [1]. Postoperative AKI can prolong the hospitalization period and increase the risk of both in-hospital mortality and chronic kidney disease [2]. In our project we propose to implement one shot learning approach with Triple Siamese Network to predict postoperative AKI after liver cancer resection.

Motivation:

AKI is a common side effect observed in patients after liver resection. Hence early prediction of postoperative AKI would help the patients as well as medical personnel to take appropriate action to reduce mortality rate. Most of the existing prediction models are based on machine learning or singular deep learning models. These models often require large balanced datasets to make good predictions. Large amount of publicly available medical datasets is hard to find, additionally they are usually imbalanced datasets. In order to handle imbalanced and small-sized datasets we propose one shot learning model with triplet loss function, first implemented in FaceNet Paper [3]. These models are widely used for image similarity detection using anchor, positive and negative triplets. We propose on applying similar approach on AKI textual data.

Datasets:

Original Dataset : Comparison of acute kidney injury between open and laparoscopic liver resection: Propensity score analysis.

The data is collected under various categories such as patient/demographics, intraoperative, laboratory and postoperative. Some of the data columns are as below:

Patient/demographics	age, sex, body mass index (BMI), diabetes
Intraoperative	crystalloid, surgery duration, platelet concentration
Laboratory	hemoglobin, platelets, creatinine, white blood cell (WBC) count, glucose and total cholesterol
Postoperative	hospital stay, vasopressor, KDIGO

The target column is Kidney Disease: Improving Global Outcomes (KDIGO) which defines postoperative AKI as an increase in serum creatinine ≥0.3 mg/dL within 2 days after surgery, or an increase ≥1.5-fold in serum creatinine within 7 days after surgery [4] .

Additional Datasets:

Prediction of Acute Kidney Injury (AKI) requires data from health care domain and this data mainly constitutes information collected from hospitals in which the patients undergo treatment. Personal information of these patients cannot be disclosed publicly and if any access has been granted, it must be treated with utmost respect and care.

We have tried to identify additional datasets which can be used in our project to predict AKI more accurately. We have found that most datasets require permission to access such datasets. Publicly accessible datasets require completion of additional training to be completed so that the information is handled appropriately. [5]

We have also found that AKI can occur as a result to many other types of conditions and surgeries [6]. Patients undergoing critical surgeries are observed in an Intensive Care Unit (ICU) of hospitals [7]. Different types of clinical data are recorded in ICU and this data can be utilized in different types of predictions including AKI. As this data is more generic as it involves different types of patient conditions, a subset of this data can be used for our project. Publicly accessible databases are available such as MIMIC [5] and eICU [8]. However, as this database contains a vast amount of patient data, appropriate mechanisms such as database scripts are required to retrieve this data [9] [10].

Additional datasets under consideration: Acute kidney injury in MIMIC, MIMIC Database, eICU database. As a future enhancement, if the required access is received for these additional datasets, we intend to use them for our project.

Background:

Acute kidney injury (AKI), a condition characterized by persistent oliguria and elevated serum creatinine levels, is a common complication in patients undergoing surgery [11]. Although the incidence of postoperative AKI depends greatly on the type of surgery, overall, 40% of in-hospital AKI cases are related to surgical procedures [12].

Machine Learning techniques have been used to predict the occurrence of AKI [2]. Our reference research paper explores machine learning methods to predict the likelihood of acute kidney injury after liver cancer resection [2]. Studies to compare performance of machine learning techniques with logistic regression models have also been conducted and in comparison of seven machine learning approaches with logistic regression analysis, the gradient boosting machine showed the best performance with the highest AUROC [13].

Convolutional Neural Network (CNN) has also been used for AKI prediction. A study aims to predict AKI by means of Convolutional Neural Network on Electronic Health Record (EHR) data [6].

System/Algorithmic design:

In the reference research paper, four machine learning algorithms- random forest, decision tree, gradient boosting (gbdt) and lightGBM (gbm) have been used [2].

We intend to implement the models with highest accuracy- random forest and gbm models as described in the reference paper [2]. In machine learning, a random forest (forest) is a classifier that includes multiple decision trees. The categories of its output are determined by the modes of categories output by individual trees. The LightGBM (gbm) algorithm is a lifting machine learning algorithm. It is a fast, distributed and high-performing gradient lifting framework based on a decision tree algorithm. It can sort, classify, run regressions, and perform many other machine learning tasks [2].

Deep learning models are used to solve complex problems in image classification, natural language processing, prediction…etc. [14]. These models require large amounts of data and many iterations to train the model. Siamese Network is a structure for non-linear metric learning which naturally learns expressions by two identical subnetworks [14]. One-shot learning technique uses related data to learn a meaningful distance metric over the space of possible inputs. This sophisticated metric is used to compare new data points to the limited available data and subsequently predict properties of these new data points [15].

The dataset available for AKI prediction has limited access. Therefore, we consider techniques that can use smaller datasets to train the model. We intend to focus our project on using one-shot learning technique and more specifically on Triple Siamese Network architecture [16] for AKI prediction.

We propose to implement two variants of Triple Siamese Networks
1) Without Transfer Learning: Convolution Neural Network (CNN) used as shared weights channel
2) With Transfer Learning: Using VGG16/MobileNetV2 as shared weights channel

Both these models will be trained on triplet loss function [17].

Evaluation plan:

The dataset would be split into train and testing datasets and testing dataset would be used to measure the performance of the model. Since it is a classification problem, we use precision, recall, F-measure, confusion matrix and ROC curves to evaluate model performance.

Task Division:

Tasks	Members
Data Analysis – Correlation heatmaps and scatter plots	Harshitha & Gargi
Feature Importance Analysis	Gargi
Data Preprocessing – Normalization & Encoding	Harshitha & Gargi
Gradient Boosting Machine	Gargi
Random Forest	Harshitha
Selecting Triples – Random and Triplet mining method [18]	Harshitha
Implementation- Triple Siamese Network	Harshitha & Gargi

Works Cited

[1] "Prevention of perioperative acute renal failure: what works?," [Online]. Available: https://doi.org/10.1016/j.bpa.2003.09.006.

[2] "A comparative study of machine learning algorithms for predicting acute kidney injury after liver cancer resection," [Online]. Available: https://doi.org/10.7717/peerj.8583.

[3] "FaceNet: A unified embedding for face recognition and clustering," [Online]. Available: 10.1109/CVPR.2015.7298682.

[4] "Acute kidney injury in leptospirosis: the Kidney Disease Improving Global Outcomes (KDIGO) criteria and mortality. Clinical Nephrology," [Online]. Available: https://www.dustri.com/article_response_page.html?artId=14895&doi=10.5414/CN108865&L=0.

[5] "MIMIC," [Online]. Available: https://mimic.physionet.org/.

[6] "Precisely Predicting Acute Kidney Injury with Convolutional Neural Network Based on Electronic Health Record Data," [Online]. Available: arXiv:2005.13171.

[7] "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database," [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3124312/.

[8] "The eICU Collaborative Research Database, a freely available multi-center database for critical care research," [Online]. Available: https://doi.org/10.1038/sdata.2018.178.

[9] "MIMIC-III, a freely accessible critical care database," [Online]. Available: https://doi.org/10.1038/sdata.2016.35.

[10] "eICU Collaborative Research Database," [Online]. Available: https://eicu-crd.mit.edu/.

[11] "Postoperative acute kidney injury," [Online]. Available: 10.4097/kjae.2017.70.3.258.

[12] "Perioperative Acute Kidney Injury," [Online]. Available: https://doi.org/10.1053/j.ackd.2012.10.003.

[13] "Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs. Logistic Regression Model," [Online]. Available: https://doi.org/10.3390/jcm7110428.

[14] "Siamese Dense Neural Network for Software Defect Prediction With Small Data," [Online]. Available: 10.1109/ACCESS.2018.2889061.

[15] "Low Data Drug Discovery with One-Shot Learning," [Online]. Available: https://doi.org/10.1021/acscentsci.6b00367.

[16] "Semi-supervised Learning using Triple-Siamese Network," [Online]. Available: https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/29087/BANERJEE-THESIS-2020.pdf?sequence=1.

[17] "Wikipedia: Triplet Loss," [Online]. Available: https://en.wikipedia.org/wiki/Triplet_loss.

[18] "Triplet Loss and Online Triplet Mining in TensorFlow," [Online]. Available: https://omoindrot.github.io/triplet-loss.

[19] "A comparative study of machine learning algorithms for predicting acute kidney injury after liver cancer resection," [Online]. Available: DOI10.7717/peerj.8583.

[20] "Comparison of acute kidney injury between open and laparoscopic liver resection: Propensity score analysis.," [Online]. Available: https://doi.org/10.1371/journal.pone.0186336.