1. State the clinical problem.

Venous thromboembolism is a preventable disease that causes more than 100,000 deaths per year. 10% of medical-surgical ICU patients will develop late stage VTE while in the hospital. Although improvements in the prophylaxis and treatment of VTE have been made, the incidence of the disease is increasing because of the difficulty of diagnosis. Currently there is no useful standard of assessing the risk of patients developing deep vein thrombosis (DVT) while in the ICU. The gold standard for diagnosis is a CT with contrast or ultrasound, however both of these are time consuming and must be performed several times a week for each patient, a task that is unfeasible for already overburdened hospital staff and economically overwhelming for patients. While the Wells Criteria and Modified Wells Criteria may be useful for assessing the possibility of DVT outside ICU settings, the validity of the test is not sensitive enough for ICU settings because almost all patients in the ICU are assessed to be at risk by these standards. While some interventions such as leg compressions are not harmful to the patient, more drastic measures, such as blood thinners, may increase the risk of other complications (e.g. hemorrhage). However, if no intervention is provided, people may develop life threatening pulmonary embolisms (PEs). Therefore, we must be able to separate patients at low risk from patients who are more susceptible to life threatening PEs.

To do this, we will determine a deep learning algorithm that can accurately, based on real time

physiological data of the patient predict whether or not the patient is at risk of developing a DVT while

under the care of a physician.

2. Describe the opportunity (what data is available and evidence that can be harnessed to solve your problem).

The data available to us is minute to minute records of approximately 10,000 patients who have been admitted into the SICU at Johns Hopkins. Since the number of patients that develop DVT in the SICU is highly selective (~2%), it is possible to develop an algorithm that correlates the patient records to the likelihood of them developing a DVT.

We know for sure if a patient has gotten DVT due to the ICD 10 code provided in the data. However, the absence of an ICD 10 code does not mean a patient did not have DVT (it could be that the DVT was not

life threatening or was missed). Furthermore, we do not know exactly when the ICD 10 code was flagged, just that the patient developed DVT. Therefore, we must determine a criteria to determine which of the patients actually developed DVT during the hospital stay, and at what time, in order to make the real-time physiological data relevant. We do know the time at which a patient may have gotten a CT with contrast, however patients may get CTs for many different reasons. We can further infer that a patient may have

been diagnosed and treated for DVT if they were subsequently treated with anticoagulants (however, a patient may have DVT but a physician may deem anticoagulants too risky, for example, due to impending surgery). There are other signs of DVT such as low PaCO2 even if supplemental oxygen is provided.

Our goal is to differentiate patients who are at risk for life threatening DVT. If a patient is presenting

with enough symptoms to be worrisome, a physician will order a CT with contrast. We can sort through

patients who have received a CT scan to determine what separates patients who were close enough symptomatically to be considered for life threatening DVT but did not actually have DVT (true negative),

from those that actually did (true positive). We cannot look outside this criterion because we do not

know if other patients may have been a false negative due to ICD 10 filtering errors, missed diagnosis,

or DVT that was simply not relevant enough compared to the other complications a patient

may have been experiencing.

3. Describe your data set.

Our data sets are raw Patient Records from hopkins hospital stored in hopkins medicine SAFE desktop. Current diagnosis are conducted mostly depending on monitored physiologic parameters and patients’

report. As a result, among all data sets, we are looking to focus on Physiological data (ICU\_Physiologic, IntraOp\_Physiologic), Vital and lab data sets as training sets for our machine learning algorithm.

Data sets are minute-to-minute, which is convenient to be read in and analyzed. Other kinds of data

such as the data of whether or not the patient develop DVT while admitted in the hospital can be used

as additional vectors to improve the accuracy of the training model.

4. Describe how you will use data to solve problems (e.g. statistical model to predict X from Y).

The current way to solve the problem is to separate the dataset into two categories; patients that develop DVT while admitted and those that did not. Next step is to develop a deep learning algorithm that can

predict whether or not a patient will develop DVT based upon the physiological data of the patient.

5. Describe how your plan breaks down into 2-3 specific aims.

Aims:

1. Categorize data into classes: Positive/Negative for VTE after ICU admission
   1. Develop a rigorous definition for VTE occurrence in our data set
   2. Investigate prevalence of VTE diagnosis in data set
2. Develop a deep learning algorithm that can predict patient risk for VTE
3. Test algorithm on an unclassified data set to determine predictive utility in clinical scenarios