Design Document: Predicting Adverse Pregnancy Outcomes with Machine Learning

Sarah McDougall
Department of Computer Science
Tufts University
Medford, MA
Sarah.McDougall@tufts.edu

I. Introduction

Maternal and infant health outcomes have become a critical area of focus within healthcare research, for maternal morbidity and mortality (MMM) and birth outcomes are an indicator of population health. [19] The recently passed Preventing Maternal Deaths Reauthorization Act and similar legislative initiatives highlight the importance in addressing maternal mortality rates by financially assisting efforts to understand pregnancy-related deaths and potential preventive measures. [2]

Despite reports estimating 60,000 women suffering from near-fatal outcomes each year, the low percentage of incidences of medical complications in the pregnant population has led to a lack of easily accessible data for analysis. [19] In addition to the scarcity of data is the presence of interrelated factors that makes adverse outcomes difficult to diagnose; this includes social determinants, medical history, genetic and immunological predisposition to certain conditions, and even environmental factors like surges in COVID-19 cases. [9] Some of these factors alter the effectiveness of certain interventions and medications due to potential effects on the embryo and fetus, and so established treatments may not be appropriate for a pregnant patient. [19] These factors prompt the need for further insight into both predictions and treatment methods that will allow clinicians to give individualized treatment to each patient. Machine learning has the ability to evaluate high-dimensional data and determine complex relationships between features and outcomes, giving the potential to increase predictive accuracy of these adverse outcomes.

This project intends to use machine learning on open-source health data to advance the health community's understanding of risk factors and indicators of adverse pregnancy outcomes for pregnant patients. While research has shown that the first trimester of pregnancy is best suited for complication prevention, this area of healthcare lacks early intervention strategies and therefore identifies conditions primarily in the third trimester of pregnancy. [21] Additionally, related studies and clinical efforts have used subjective analyses for diagnosis and treatment, and so machine learning can support clinicians in their clinical decisions, ensuring that patients receive appropriate diagnoses and treatment.

This design document describes the multimodal, multi-task machine learning model that will be built to predict adverse pregnancy outcomes on the open-source Multiparameter Intelligent Monitoring in Intensive Care (MIMIC)-IV dataset. In doing so, the model will aim to determine the most important features for prediction and determine novel risk factors that are not traditionally monitored in clinical risk assessments.

II. Hypotheses

This project will explore the following three hypotheses:

1) Certain clinical features exhibit a significant association with adverse pregnancy outcomes, suggesting that they should be monitored in clinical practices as a means of early detection and prevention.
2) The constructed model will suggest features of importance that align with existing research efforts in this field. Additionally, the model may suggest novel risk factors not traditionally monitored in clinical risk assessments, such as demographic and socioeconomic factors.

3) The constructed model will detect a pattern in predictive accuracy across the first, second, and third trimesters of pregnancy.

III. Background

The following sections describe the three conditions of interest and offer key insights into adverse pregnancy outcomes from related literature.

A. Preeclampsia

Preeclampsia is both a common and severe complication in pregnancy that is characterized by a range of symptoms, including hypertension (high blood pressure), high protein levels in urine that indicate kidney damage, and other symptoms of maternal end-organ damage, including damage to the central nervous system, lungs, liver, kidneys, and heart. Preeclampsia in pregnancy may present itself as either early- or late-onset preeclampsia. It has been hypothesized that early-onset preeclampsia begins with abnormal implantation and placentation in the first trimester of pregnancy, whereas late-onset preeclampsia is typically diagnosed after 20 weeks of gestation and is due to abnormalities in blood flow to the placenta. Late-onset preeclampsia may be a result of maternal vascular disease, which is often associated with patients who have conditions such as chronic hypertension and pre-gestational diabetes. [17] Studies suggest that the development of preeclampsia may begin in the first trimester, although its symptoms occur most often in the third trimester. [9] [27]

Because the exact cause of preeclampsia has not been fully understood, and because women with preeclampsia have a 20-time higher maternal mortality rate, it is important to study the heterogeneity of how preeclampsia presents itself throughout pregnancy. Related studies have indicated that predicting adverse outcomes via machine learning will help clinicians to more appropriately predict the timing of disease onset and develop the appropriate treatments for patients. [21]

B. Premature Delivery

Premature delivery, synonymous with preterm birth, refers to deliveries that occur before 37 weeks of gestation. Premature birth can range significantly in severity, with 34-36 weeks considered late preterm and before 28 weeks of gestation considered extremely preterm. [5] Prematurity is considered the leading cause of infant morbidity and mortality. [9] While this condition may cause serious health problems for the infant, it also may cause physical and emotional issues for the mother as well. Studies have shown that mothers of premature infants may experience negativity and are at a higher risk of meeting the criteria for clinical depression. [7] [11]

C. Obstetric Hemorrhage

An obstetric hemorrhage is characterized by excessive bleeding that may occur at different points in time, including before childbirth (antepartum), during childbirth (intrapartum), and afterward (postpartum). Over 80 percent of cases occur postpartum. While there are some identifiable risk factors such as anemia and underlying bleeding disorders, most women who experience an obstetric hemorrhage will not have any identifiable risk factors. Despite the often lack of identifiable risk factors, the primary method of prevention is through reassessment of risk factors throughout pregnancy to prevent a potential hemorrhage from becoming life-threatening. Many of the deaths that occur as a result of hemorrhage are due to a poor clinical response, which prompts the need for further research into risk factors and methods for prediction. [16]

D. Other Conditions

Among the most commonly reported adverse pregnancy outcomes are preeclampsia and peripartum hemorrhage. [3] However, there are several other complications that need further research into their causes, including fetal growth restriction, stillbirth, and sepsis. Maternal adverse outcomes may also be defined to encompass the following conditions: cerebral hemorrhage, disseminated intravascular coagulopathy (DIC), pulmonary edema, renal failure, hemolysis, elevated liver enzymes, low platelet count, eclampsia, and death.

E. Related Works

From the small number of published studies related to predicting adverse maternal health outcomes are results that suggest a high heterogeneity between control groups and groups with adverse effects. While these published studies reveal potential variables that are important for prediction, these studies also exhibit key limitations in terms of modality and size of the data used for analysis. One related study analyzed the relationship between maternal preeclampsia and the development of respiratory distress syndrome (RDS) among infants with very low birth weight (VLBW). The study's results contradicted previous findings that preeclampsia is considered a protective factor against RDS by accelerating fetal lung maturation. Instead, this study found that early-onset preeclampsia increased the risk of severe RDS, while a previous study found a decreased risk when examining the outcomes of late-onset preeclampsia. This study highlights the many confounding factors that may be present when analyzing maternal adverse outcomes. [25]

Another study used a machine learning framework to identify severe maternal morbidity (SMM) and relevant risk factors from EHR data. Several logistic regression models were built with combinations of diagnosis and procedure data fields from the EHRs. The study reveals several important concepts for prediction, including: dependence on ventilator, intubation, critical care, acidosis, and sepsis. While the study is well-designed, the analysis is limited to the use of one of the simplest machine learning algorithms and only includes diagnosis and procedural features. [10]

A 2021 systematic review demonstrated the sparsity of publications in its analysis of 31 suitable articles after excluding those that did not meet the search criteria. Of these publications, seven were related to prematurity, while six were related to preeclampsia. Of the 21 different machine learning methods used across the 31 articles are the AdaBoost model for targeting preeclampsia and Support Vector Machine (SVM) for targeting prematurity. [17] [20] In general, prematurity prediction outperformed pre-eclampsia in terms of the AUC value obtained from performance evaluation across the studies. Only four studies were found to use deep learning methods; despite the low number of publications, they were all considered to have good predictive performance. Among these four studies is one that targets preeclampsia using biological markers with a multilayer perceptron. [18] This systematic review reveals that diverse efforts have been made toward predicting adverse outcomes, as demonstrated by the variety of machine learning models, modality of data, and target variables. The review reveals important predictive variables that can be used for comparison of results across models, including systolic blood pressure, serum levels, platelet count, and blood glucose level for preeclampsia. [4]

IV. Study Data

The Multiparameter Intelligent Monitoring in Intensive Care (MIMIC)-IV dataset is a publicly available database sourced from the electronic health record (EHR) of the Beth Israel Deaconess Medical Center (BIDMC) in Massachusetts. The dataset contains data on a clinical cohort of patients that were admitted to the Emergency Department (ED) or an intensive care unit (ICU) between the years of 2008 and 2019. All patients are greater than 18 years of age and the patient records have been de-identified to abide by HIPAA regulations.

The MIMIC-IV dataset takes on a relational structure and its data is organized into several modules that are depicted by the source of the data. Each module consists of multiple tables that capture the different

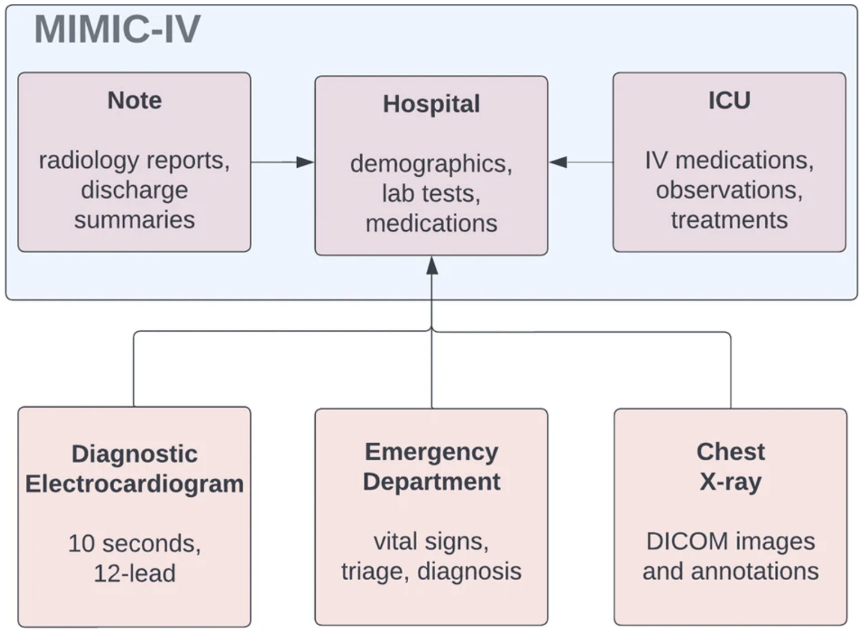

Fig. 1: Diagram of the relational structure of the MIMIC-IV modules. [12]

aspects of clinical admissions and observations. The tables can be easily linked via common identifiers, such as a "subject" ID or "admission" ID. The three main modules included in MIMIC-IV are as follows:

1) The "Hospital" module - Contains data acquired from the hospital-wide EHR. Includes billed events, medication prescriptions and administration, microbiology cultures, and admission/discharge/transfer (ADT) records.
2) The "ICU" module - Contains data gathered from the clinical information system used within the ICU. Includes IV infusions, patient outputs, and charted observations during the stay.
3) The "Note" module - Contains free-text notes for both discharge summaries and radiology reports.

In addition to the data gathered internally, the MIMIC-IV dataset also contains non-clinical external data sources that are used to better understand the collected data. This includes a mapping table with the International Classification of Diseases (ICD) codes, Diagnosis Related Groups (DRGs), and the Healthcare Common Procedure Coding System (HCPCS).

This dataset was chosen for analysis for several reasons. First, most other openly accessible datasets only contain one modality of clinical information, most notably clinical observations. The MIMIC-IV dataset, by contrast, contains imaging, free-text and genomic data in addition to hospital admission and ICU data. Second, as noted in related studies, most analyses are based on very small sample sizes. The MIMIC-IV dataset contains approximately 180,000 unique subjects in its admission module alone, and approximately 12,000 of those subjects contain a pregnancy diagnosis (as defined by the pregnancy ICD-9 and ICD-10 codes). This is a sufficient sample size, especially in comparison to related studies.

V. Methods

The following sections outline the proposed methods for this project, including technologies used, data pre-processing methods, the model architecture, and performance evaluation metrics.

A. Technologies Used

Programming language: Python
Libraries and Frameworks: TensorFlow, scikit-learn, keras, pandas, numpy
Development Environment: Jupyter Notebook
Cloud Platform: Google Cloud Platform (GCP) (Google BigQuery)

B. Output Data Architecture

The constructed model will aim to predict three distinct adverse pregnancy outcomes: (1) preeclampsia (occurring during any trimester), (2) preterm delivery, and (3) obstetric hemorrhage (including postpartum hemorrhage).

The output data will comprise three binary columns, with each column representing one of the three outcomes. Each of these columns will be defined by whether the patient was billed a diagnosis that falls under the predetermined list of ICD-9/ICD-10 codes for the given condition. Since the ICD-9 revision was replaced in 2015 by the ICD-10 revision, both revisions will be used in conjunction with each other to avoid any loss of information in translating between versions. Note that a patient may have several "positive" outcomes and thus may have multiple outcome columns with the same value.

Within the MIMIC-IV dataset, billed diagnoses are associated with an admission instance. Since each admission has a recorded timestamp, we will use these timestamps for constructing the output columns. Generally, the adverse outcome diagnosis must have occurred after a pregnancy diagnosis was recorded, and also be within one year of the pregnancy diagnosis being recorded. This restriction prevents instances from being flagged whose diagnoses are related to a subsequent pregnancy or may not be tied to a pregnancy.

The figure below outlines a subset of the diagnosis codes that will be used for defining the binary output columns. Note that the ICD-10 code collection O00-O9A relates to pregnancy, childbirth, and puerperium, and so most ICD-10 codes used for this analysis will belong to this collection. Within this collection are sub-collections for "Edema, proteinuria, and hypertensive disorders in pregnancy, childbirth, and the puerperium" (O10-O16), "Other maternal disorders predominantly related to pregnancy" (O20-O29), "Maternal care related to the fetus and amniotic cavity and possible delivery problems" (O30-O48), among others. A condition may be encompassed by one or more of these sub-collections.

Fig. 2: Example mapping of target output conditions to corresponding ICD-10 collections or code(s).

Condition	ICD-10 Collection	Details
Preeclampsia	O14	Includes mild to moderate preeclampsia and severe preeclampsia. Includes second trimester, third trimester, and unspecified trimester diagnoses.
Preterm Delivery	O60.1	O60 includes preterm labor without delivery and term delivery with preterm labor.
Obstetric Hemorrhage	O20; O72	O20 represents hemorrhage in early pregnancy. O72 represents post-partum hemorrhage, including third-stage hemorrhage and secondary postpartum hemorrhage.

C. Input Data Architecture

The input data will consist of features derived from the MIMIC-IV dataset. The three target outcomes will share a single set of input features. While the input data contains some static features, such as patient demographics attributes and social determinants, the dataset primarily contains temporal data. Examples of temporal data within the MIMIC-IV dataset include: unstructured clinical notes, ICU data for patient monitoring, and measures of vital signs. In general, EHR records are structured so that they are a general span over several years in order to capture patient trajectories and outcomes over time.

D. Data Preprocessing

The inclusion criteria for this study is as follows: the patient must be at least 18 years of age with a recorded pregnancy diagnosis. As part of the data pre-processing pipeline, the eligible patient population will be filtered according to the collection of ICD-9 and ICD-10 diagnosis codes for pregnancy diagnosis codes. Many studies have suggested that structured data searches may not retrieve all relevant patients. Therefore, we will additionally filter the data for the eligible patient population via the unstructured

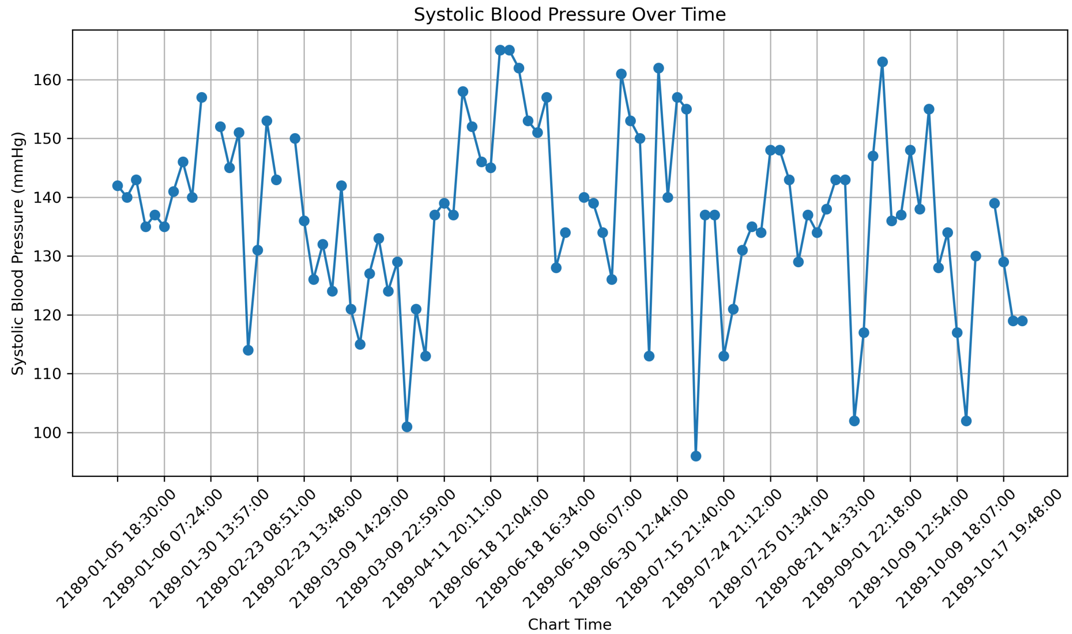

Fig. 3: Example graph illustrating the sequential nature of the vital signs data. The graph plots systolic blood pressure readings over the course of a de-identified year for a single patient who has a recorded pregnancy diagnosis.

note text with search terms such as “pregnancy” and “pregnant.” [8] After retrieving the initial patient population, we will tag patients who have ICD-9 or ICD-10 diagnoses for the following conditions: (1) preeclampsia, (2) preterm delivery, and (3) obstetric hemorrhage. This will create our target columns for the analysis and define which patient records belong to the target classes.

In addition to filtering the initial dataset to the relevant patient population, the unstructured note data contained in the MIMIC-IV dataset will also need to be pre-processed for further analysis. One route for preprocessing that will be attempted is to use the scispaCy Python package. This package is similar to the Natural Language Toolkit (NLTK) Python package in that it contains a tokenizer, a part-of-speech (POS) tagger, and other text preprocessing functionalities. The sciSpacy package has several available models that can be used, including a model of approximately 785,000 biomedical vocabulary terms. With this package, it may be possible to conduct named entity recognition (NER) and classify the unstructured note data into predefined medical categories such as diseases and diagnoses, symptoms, medications, and procedures. [1]

E. Feature Selection

In the related works that were analyzed, the number of features included in machine learning and deep learning models ranged from 30 to nearly 3000 features. The approach for identifying the most informative features prior to feeding the data into the model will rely on both (1) domain knowledge from related works, and (2) an analytical approach to remove features that are deemed redundant in terms of the information they provide to the model.

1) Using Domain Knowledge from Related Works: Several publications have built models to predict select adverse pregnancy outcomes like preeclampsia. These works, combined with domain knowledge from clinical practices, can influence the features that are included in the final model for this experiment. In general, hematology panels and urinalysis provide useful information about maternal and fetal health. In particular, urine cultures determine proteineuria (abnormal amount of protein in urine) and bactueriuria (presence of bacteria in urine), which are known predictions of conditions like preeclampsia. [13] [22]

Additionally, it has been found that higher neutrophil (a type of white blood cell) counts and neutrophil-to-lymphocyte ratios, both inflammatory markers, are associated with decreased fetal growth and thus may be useful for our model. [15] High-level indicators of adverse outcomes include micronutrient deficiencies, specifically vitamin D, malnutrition, obesity, and high stress levels. [9]

2) Removing Redundant Information: T. Nair's analysis of preeclampsia using microarray data, which records expression values of thousands of genes simultaneously, involved a statistical feature selection process that may be beneficial in the context of this project. [18] The feature selection process used a statistical multiple comparison-based method that is similar to an ANOVA F-Test while additionally providing information about which means are different rather than simply stating whether the difference in means is considered significant. By discerning which means are different, this approach can identify features that vary in their presentation across different target groups. Conversely, this approach can identify features that present similarly across the target groups, signaling that these features are considered redundant and can be removed from the model without removing any key information. Another study used a random forest model to select the most important features from ICU observation data, based on the highest classification accuracy. This feature selection approach could also be applied to this project, although random forest models have the potential to overfit the training data when working with a large number of features and are computationally expensive. [26]

F. The Multi-Task Learning Architecture

As noted in previous sections, this project will involve working with high-dimensional data and will model complex relationships among clinical features and their associated outcomes. Because each input record is associated with clearly defined output labels, this project will use a supervised learning algorithm for predicting adverse pregnancy outcomes. This project will build an artificial neural network (ANN) for analyzing the filtered MIMIC-IV dataset since the input data includes complex data formats like time-series measurements and free-text notes.

The constructed model will be built using the TensorFlow framework and its architecture will be built for a multi-task learning environment. Healthcare is one of many domains where multi-task learning can be beneficial, as predicting several clinical outcomes in patients at the same time allows for the model to find associations between different adverse outcomes and improve overall diagnostic accuracy by exploiting useful information from one given task to another. [24] Specifically with multi-modal multi-task learning, introducing another layer of obstruction may yield even more predictive benefits than single-modal multi-task learning. [6]

Multi-task learning is closely related to transfer learning, for transfer learning is a special case of multi-task learning. While in transfer learning there is an assumption that a secondary task can provide additional context about the primary task being analyzed, multi-task learning instead treats all tasks equally. [23] Multi-task learning is thought to relate closely to human learning in that the concepts that are learned are generalizable across different settings, contributing to the learning of relationships and overall faster learning. [6]

Multi-task models can vary in terms of their input/output structure and their means of parameter sharing. Input/output variations include: (1) multi-input single-output, (2) single-input multi-output, and (3) multi-input multi-output. This project will utilize a single-input multi-output model since the tasks will share the same set of input data rather than using data from different sources. All three tasks for our model will be using the same features from the MIMIC-IV dataset. In terms of parameter sharing, the built model will likely use hard parameter sharing, meaning that the hidden layers will be shared between all tasks, as opposed to soft parameter sharing where each task has its own hidden layers with their own parameters. [23] Note that there exist several variations of parameter sharing beyond these two broad categories, as outlined in [6].

The nature in which information is shared among the output tasks depends on the relationship of the tasks, which can be learned from the data or established beforehand. For the tasks selected for our model,

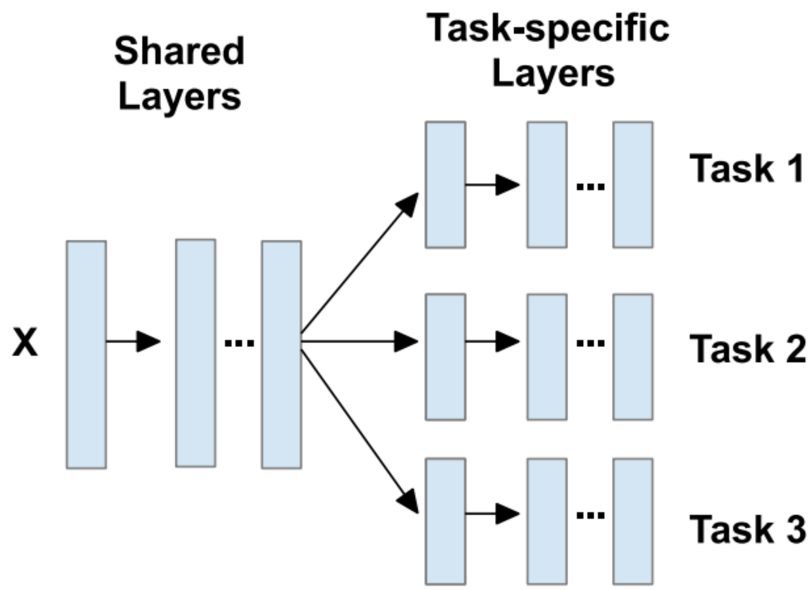

Fig. 4: Multi-task learning architecture for deep learning. [23]

it can be assumed that the relationship is not hierarchical, although other relationship structures may exist and may be learned from the data. Task relationship learning (TRL) is a concept specific to multi-task learning with the goal of discerning relationships between tasks. This may be further explored throughout model construction and training.

Note also that the constructed model will need to strike a balance between the degree of information shared between tasks without causing negative transfer to occur. It is possible that increasing the performance of one task may hinder performance of another task because the two tasks may have different needs. A known difficulty of multi-task learning is optimizing for multiple loss functions, and so this is an area of research that will require more investigation when constructing the model. One possible approach that has been used is to combine many loss functions into one using a weighted average, while another approach is to use multi-objective optimization and optimize several objective functions simultaneously. These factors will be considered when constructing the model. [6]

The TensorFlow framework can be used to construct the model architecture described above. The TensorFlow model can be built so that some parts of the network are shared among tasks while others are specific to a given task. The model can employ joint training in which the tasks are trained at the same time on the single dataset. The parameters of the shared layers can be updated during the training process from each task. [14]

G. Performance Metrics

The following performance metrics will be calculated and evaluated for the constructed model:

1) **Area Under the Receiver Operating Characteristic Curve (AUC)**. This metric measures the ability of the model to differentiate between the positive and negative classes across different thresholds. This metric is commonly used in publications within this domain, and so it will be beneficial to calculate and compare this metric with published machine learning models.
2) **Recall (sensitivity)**. Recall measures the ratio of true positive predictions to the sum of true positives and false negatives. We aim for the model to avoid misclassifications, especially false negatives. Recall focuses on minimizing false negatives and maximizing positive instances.

Other metrics of importance include false positive rate, false negative rate, and the precision-recall curve. It will also be beneficial to construct confusion matrices to visualize the false positive and false negative rates of the model. The above metrics are suitable for this project because they are appropriate for measuring the performance of imbalanced datasets, unlike other performance metrics like accuracy.

H. Addressing Bias

As with many other areas of research within the clinical domain, the dataset used for this project will be "imbalanced" in that the proportion of patients that experience adverse outcomes will be significantly smaller than the proportion of pregnant patients that do not experience any of the adverse outcomes being evaluated. An initial investigation into the MIMIC-IV dataset revealed a total of 12,000 unique patients with a pregnancy diagnosis, with only 200 unique patients with a preeclampsia diagnosis occurring after the initial pregnancy diagnosis. These initial results reveal a significant imbalance among the class labels. Therefore, techniques will need to be employed to ensure that the trained model avoids bias as much as possible.

Fortunately, several techniques exist for addressing imbalanced datasets. One such method is class weighting, in which misclassifications of the minority class can be penalized more heavily during training. TensorFlow's loss functions support class weighting, and so this can be accounted for using this built-in mechanism within the algorithm. Resampling techniques like the Synthetic Minority Over-sampling Technique (SMOTE) can be employed to improve model performance and address the class imbalance. SMOTE creates synthetic samples for the minority class using a "nearest neighbors" approach so the minority class is more heavily represented, reducing overfitting of the training model.

I. Stretch Objectives

If time permits, this project will explore alternative approaches for analyzing the unstructured note text. A related study analyzed the discharge summary data on the MIMIC II/III datasets using a concept-based search engine tool called "Essie." The engine uses a structured representation of medical concepts and rules that go beyond simply matching keywords. Using such a tool to search clinical notes for pregnancy could increase the number of patients included in the initial patient population. The "Brat Rapid Annotation Tool" (BRAT) could be used in conjunction with the Essie search engine to manually extract certain categories of information from the notes, including patient demographics, medications, diagnoses, and past medical history. It may be beneficial to explore this open-source browser-based tool as an alternative tool for conducting named entity recognition. [19] These additional tools may improve the performance metrics when used to build the model.

The project design can be extended by other means if time permits. One area of interest that can be further explored is hyperparameter optimization, specifically using Bayesian approaches. When determining the next set of hyperparameters for evaluation, Bayesian optimization considers previously evaluated results. Bayesian optimization and other hyperparameter approaches have the potential to enhance the final model and offer new insights into how the model behaves.

Lastly, other models can be constructed to give a fuller understanding of how adverse pregnancy outcomes are presented by observing outcomes that go beyond the three trimesters of pregnancy. Models can be constructed that observe postpartum outcomes, as the postpartum period is often overlooked in literature despite being very important for the mother's physical and emotional health.

J. Backup Plan

The proposed project design is ambitious because of the small amount of published clinical applications of multi-task learning and small amount of deep learning efforts on clinical data as it relates to pregnancy. There are several alternative approaches and modifications that can be made to the project design in the case where the initial design cannot be completed in the 12-week timeframe. These modifications are outlined below.

1) Modify the multi-task output architecture. Using three targets in the model may be difficult to construct programmatically or may give ambiguous results. If this is the case, the number of targets can be lessened to one or two outputs. The modified output may be one of the target outputs defined in previous sections (for example, preeclampsia), or the output may consist of a catch-all "adverse pregnancy outcome" target, as some related works have used in their models. Using an "adverse pregnancy outcome" target may be appropriate if it successfully mitigates bias that would otherwise be present under a different approach.
2) This project uses multi-modal data, including time-series data. In the case where it is deemed unfeasible to incorporate the time-series data into a deep learning model, tabular data can be extracted from the MIMIC-IV dataset and used to construct a machine learning model that can handle tabular data. Several works have employed logistic regression, AdaBoost, and Support Vector Machine (SVM) models using tabular clinical data and they have achieved moderate success in predicting adverse pregnancy outcomes.
3) The text data analysis on discharge summaries may be omitted from the project design if the anticipated work cannot be fully completed in the allotted 12-week timeframe.

VI. Project Timeline

The following table outlines a high-level breakdown of the project timeline.

TABLE I: Project Timeline

Week Of	Tasks
May 20	Set up project development environment, review documentation for relevant tooling
May 27	Set up infrastructure for housing the data
June 3	Data preprocessing on non-text data
June 10	Data preprocessing on text data using relevant Python packages
June 17	Exploratory data analysis and feature engineering
June 24	Multi-task learning research, build ANN model
July 1	Build ANN model
July 8	Train initial models and evaluate performance
July 15	Analyze model performance, iterate on model training
July 22	Fine-tune model, create performance visualizations
July 29	Summarize findings
August 5-13	Presentation preparation, delivery, and documentation

References

[1] Allen Institute for AI. SciSpaCy: A Python package for NLP in biomedical research. SciSpaCy website, n.d.

[2] American Hospital Association. House Passes AHA-Supported Preventing Maternal Deaths Reauthorization Act. AHA News, March 6 2024.

[3] James Bale, Barbara Stoll, and Alan Lucas. *Improving Birth Outcomes: Meeting the Challenge in the Developing World*. National Academic Press, Washington, DC, 1st edition, 2003.

[4] Andrea Bertini, Rodrigo Salas, Stéphanie Chabert, Luis Sobrevia, and Fabián Pardo. Using machine learning to predict complications in pregnancy: A systematic review. *Front Bioeng Biotechnol*, 9:780389, Jan 19 2022.

[5] Mayo Clinic. Premature Birth: Symptoms & Causes. Mayo Clinic website, n.d.

[6] Michael Crawshaw. Multi-task learning with deep neural networks: A survey, 2020.

[7] Leigh Davis, Helen Edwards, Heather Mohay, and Judy Wollin. The impact of very premature birth on the psychological health of mothers. *Early Human Development*, 73(1):61–70, 2003.

[8] Travis Edinger, Dina Demner-Fushman, Aaron M. Cohen, Steven Bedrick, and William Hersh. Evaluation of clinical text segmentation to facilitate cohort retrieval. *AMIA Annu Symp Proc*, 2017:660–669, Apr 16 2018.

[9] Camilo Espinosa, Martin Becker, Ivana Marić, Ronald J. Wong, Gary M. Shaw, Brice Gaudilliere, Nima Aghaeepour, David K. Stevenson, Ina A. Stelzer, Laura S. Peterson, Alan L. Chang, Maria Xenochristou, Thanaphong Phongpreecha, Davide De Francesco, Michael Katz, Yair J. Blumenfeld, and Martin S. Angst. Data-driven modeling of pregnancy-related complications. *Trends in Molecular Medicine*, 27(8):762–776, 2021. Reproductive and Sexual Health.

[10] Cheng Gao, Sarah Osmundson, Xia Yan, David V. Edwards, Bradley A. Malin, and Yunan Chen. Learning to identify severe maternal morbidity from electronic health records. *Stud Health Technol Inform*, 264:143–147, Aug 21 2019.

[11] Chiara Ionio, Camilla Colombo, Valeria Brazzoduro, Eleonora Mascheroni, Emanuela Confalonieri, Francesca Castoldi, and Gianluca Lista. Mothers and fathers in nicu: The impact of preterm birth on parental distress. *Eur J Psychol*, 12(4):604–621, Nov 18 2016.

[12] Alistair E.W. Johnson, Lorenzo Bulgarelli, Linda Shen, and et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*, 10:1, 2023.

[13] J. Kaduma and et al. Urinary tract infections and pre-eclampsia among pregnant women attending two hospitals in mwanza city, tanzania: a 1:2 matched case-control study. *Biomed. Res. Int.*, 2019, 2019.

[14] KDnuggets. Multi-Task Learning in TensorFlow: Part 1. KDnuggets website, 2016.

[15] O. Levy and et al. Early predictors of small-for-gestational-age neonates using non-invasive, low-cost, and readily available hematological markers. *Int. J. Gynaecol. Obstet.*, 150:340–345, 2020.

[16] C. McLintock and A.H. James. Obstetric hemorrhage. *Journal of Thrombosis and Haemostasis*, 9(8):1441–1451, 2011.

[17] Sarah Munchel, Suzanne Rohrback, Carlo Randise-Hinchliff, Sarah Kinnings, Shweta Deshmukh, Nagesh Alla, Catherine Tan, Amirali Kia, Grainger Greene, Linda Leety, Matthew Rhoa, Scott Yeats, Matthew Saul, Julia Chou, Kimberley Bianco, Kevin O'Shea, Emmanuel Bujold, Errol Norwitz, Ronald Wapner, George Saade, and Fiona Kaper. Circulating transcripts in maternal blood reflect a molecular signature of early-onset preeclampsia. *Science Translational Medicine*, 12(550):eaaz0131, 2020.

[18] T.M. Nair. Statistical and artificial neural network-based analysis to understand complexity and heterogeneity in preeclampsia. *Comput Biol Chem*, 75:222–230, Aug 2018.

[19] Laura M. Rodriguez and David D. Fushman. Finding understudied disorders potentially associated with maternal morbidity and mortality. *AJP Rep.*, 9(1):e36–e43, Jan 2019.

[20] Nada Sadi-Ahmed, Benamar Kacha, Hamid Taleb, and et al. Relevant features selection for automatic prediction of preterm deliveries from pregnancy electrohysterographic (ehg) records. *J Med Syst*, 41:204, 2017.

[21] Leon J. Schmidt, Oliver Rieger, Mark Neznansky, Max Hackelöer, Lisa A. Dröge, Wolfgang Henrich, David Higgins, and Stefan Verlohren. A machine-learning–based algorithm improves prediction of preeclampsia-associated adverse outcomes. *American Journal of Obstetrics and Gynecology*, 227(1):77.e1–77.e30, 2022.

[22] Atalay Tanacan and et al. The importance of proteinuria in pre-eclampsia and its predictive role in maternal and neonatal outcomes. *Hypertens. Pregnancy*, 38:111–118, 2019.

[23] Kim-Han Thung and Chong-Yaw Wee. A brief review on multi-task learning. *Multimed Tools Appl*, 77:29705–29725, 2018.

[24] Willem Waegeman, Krzysztof Dembczyński, and Eyke Hüllermeier. Multi-target prediction: a unifying view on problems and methods. *Data Min Knowl Disc*, 33:293–324, 2019.

[25] Yi-Han Wen, Hsin-I Yang, Hung-Chieh Chou, and et al. Association of maternal preeclampsia with neonatal respiratory distress syndrome in very-low-birth-weight infants. *Sci Rep*, 9:13212, 2019.

[26] X. Wu, R. Li, Z. He, and et al. A value-based deep reinforcement learning model with human expertise in optimal treatment of sepsis. *npj Digit. Med.*, 6:15, 2023.

[27] Agnieszka Wójtowicz, Marta Zembala-Szczerba, Dominika Babczyk, Monika Kołodziejczyk-Pietruszka, Olga Lewaczyńska, and Hubert Huras. Early- and late-onset preeclampsia: A comprehensive cohort study of laboratory and clinical findings according to the new ishhp criteria. *Int J Hypertens*, 2019:4108271, Sep 17 2019.