Machine Learning in Health - Project Proposal

Project: Machine Learning-Based Mortality Prediction in ICU Patients with Stage II Hypertension: A Retrospective Analysis Using the MIMIC-IV Database

Members:

Wedad Amer	1081635
Jiahe Liu	1214235
Quan Yi	1054540
Vaibhav Singh Makhloga	1424591
Jay Acharya	1401435

1. What is the clinical question/problem your group will tackle and why is it relevant? Please include information about the magnitude of the problem and make sure your claim is supported by references to the literature (100 words).

Stage II hypertension, characterised as blood pressure ≥*160/100 mm Hg*, is more frequently associated with cardiocerebrovascular disease, ICU admission, and mortality (Salgado et al., 2013). ICU patients with this condition can rapidly deteriorate and require enhanced monitoring (Schweingruber et al., 2022). Previous research has primarily centred on identifying demographic risk factors for Stage II hypertension (Ostchega et al., 2022) and clinical risk factors for stroke (Huang et al., 2023). This research aims to identify demographic and clinical indicators for Stage II hypertension and will employ binary machine learning to predict mortality in these ICU patients, facilitating more precise monitoring.

2. Which data source(s)and information will you use? What is your plan for data processing and digital phenotyping? Remember for a digital phenotype it is important to list the criteria and describe which data elements you will use to find patients that meet those criteria (194 words, including table).

Using the MIMIC-IV ICU dataset, our primary focus revolves around demographic and clinical risk factors in hypertensive disease. We've correlated relevant literature (Huang et al., 2023) with MIMIC-IV data, selecting features from table 1. Further risk factor analysis will be conducted to include significant features in machine learning.

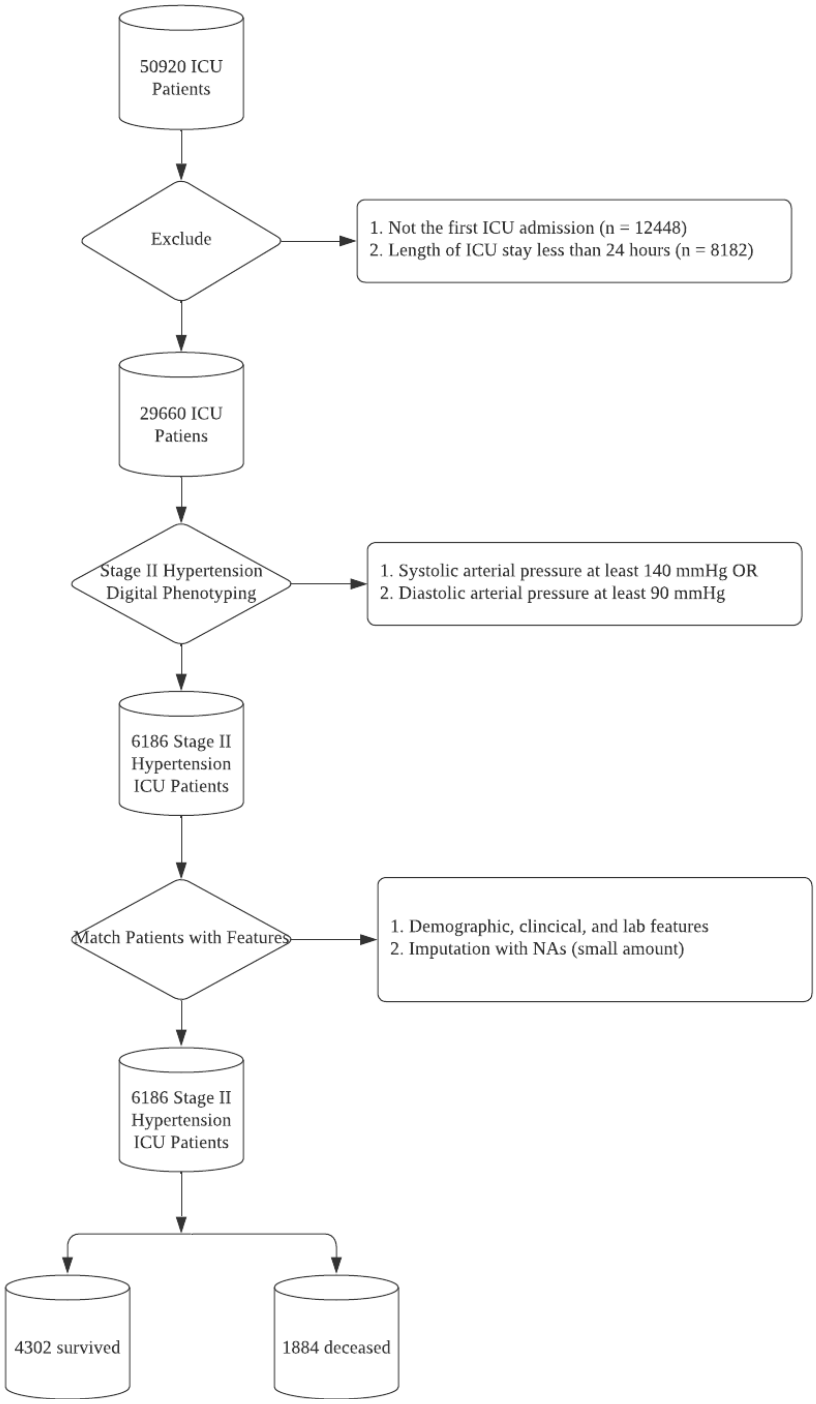
Table 1. Risk factors

	MIMIC Table	Features	itemid
Patients' Demographics	mimiciv_hosp.patients	Gender	\
		Age	\
Clinical Measurements	mimiciv_icu.d_items	Heart rate	220045
		Systolic blood pressure	220050
		Diastolic blood pressure	220051
		Respiratory rate	220210
		Heart rate Alarm - High	220046
		HrApacheIIScore	226763
		RR > 35 for > 5 min	224718
Lab Measurements	mimiciv_hosp.labitems	Glucose	50809
		Hemoglobin	50811
		Potassium, Whole Blood	50822
		Temperature	50825
		Hematocrit	52028
		Anion Gap	50868
		Bicarbonate	50882
		Chloride	50902
		Creatinine	50912
		Sodium	50983
		Urea Nitrogen	51006
		MCV	51691
		White Blood Cells	51755
		MCH	51248
		MCHC	51249
		PT	51274
		RDW	51277
		Rbc	52170

For preprocessing, we'll only consider adult patients with a single ICU stay exceeding 24 hours. This ensures synchronised, one-to-one time matching.

Our digital phenotype research targets the MIMIC-IV ICU stage II hypertension cohort, specifically identifying patients based on :

1. systolic arterial pressure [SAP] >160 mmHg OR
2. diastolic arterial pressure [DAP] >100 mmHg (Chobanian, 2003)



Tabel 1. Data Pipeline

3. What is your machine learning approach (flavour),methodology, metrics of success and expected outcomes? (300 words)

As the purpose of this project is to conduct a retrospective analysis of patients with type II hypertension and to determine their risk of mortality, it is important to consider several approaches to ensure that we generate an optimal classifier. For this project, we will be exploring supervised machine learning approaches to create a binary mortality prediction classifier (alive/deceased), given the features presented in *section 2*.

In the project, we will thoroughly explore, compare and contrast the differences between the selected models and determine whether these models are appropriate approaches for this particular clinical scenario. The specific classifiers we will be focusing on are Naive Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), k-nearest Neighbours (KNN), Random Forest, and a Deep Neural Network (DNN). The selected models will be implemented through Python's scikit-learn library, allowing us to not only efficiently explore different approaches given the derived phenotypes from MIMIC-IV, but also to optimise each of the models' prediction performance through easy hyperparameter tuning.

We will carefully adjust the selection of features based on measures such as accuracy, precision, recall, and F1-score to maximise the model performance. Given that the dataset we will conduct feature selection and handle missing values during the preprocessing of the data before feeding it into the classifier. Furthermore, given that the derived phenotype is unbalanced; 4302 alive and 1884 deceased, we will also be utilising k-fold cross-validation to create a robust and reliable model to minimise any bias and variance. This meticulous process ensures that the model comparison and evaluation will have a strong, reliable foundation. The purpose of this project is to maximise the precision and recall of the mortality predictions, and to reduce the false negative errors, gathering insights into which patients require the most attention during a type II hypertensive episode.

References

Chobanian, A. V. (2003). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood PressureThe JNC 7 Report. JAMA, 289(19), 2560. https://doi.org/10.1001/jama.289.19.2560

Huang, J., Chen, H., Deng, J., Liu, X., Shu, T., Yin, C., Duan, M., Fu, L., Wang, K., & Zeng, S. (2023). Interpretable machine learning for predicting 28-day all-cause in-hospital mortality for hypertensive ischemic or hemorrhagic stroke patients in the ICU: A multi-center retrospective cohort study with internal and external cross-validation. Frontiers in Neurology, 14, 1185447. https://doi.org/10.3389/fneur.2023.1185447

Ostchega, Y., Hughes, J. P., Kit, B., Chen, T.-C., Nwankwo, T., Commodore-Mensah, Y., Graber, J. E., & Nguyen, D. T. (2022). Differences in Hypertension and Stage II Hypertension by Demographic and Risk Factors, Obtained by Two Different Protocols in US Adults: National Health and Nutrition Examination Survey, 2017–2018. *American Journal of Hypertension*, *35*(7), 619–626. https://doi.org/10.1093/ajh/hpac042

Salgado, D. R., Silva, E., & Vincent, J.-L. (2013). Control of hypertension in the critically ill: A pathophysiological approach. *Annals of Intensive Care*, *3*(1), 17. https://doi.org/10.1186/2110-5820-3-17

Schweingruber, N., Mader, M. M. D., Wiehe, A., Röder, F., Göttsche, J., Kluge, S., Westphal, M., Czorlich, P., & Gerloff, C. (2022). A recurrent machine learning model predicts intracranial hypertension in neurointensive care patients. *Brain*, *145*(8), 2910–2919. https://doi.org/10.1093/brain/awab453