COMP90089 GROUP ASSIGNMENT

MACHINE LEARNING-BASED MORTALITY PREDICTION IN ICU PATIENTS WITH ACUTE KIDNEY INJURY: A RETROSPECTIVE ANALYSIS USING THE MIMIC-IV DATABASE

Jiahe Liu
1214235
jiahe3@student.unimelb.edu.au

Quan Yi
1054540
quyi@student.unimelb.edu.au

Wedad Amer
1081635
wamer@student.unimelb.edu.au

Vaibhav Singh Makhloga
1424591
vmakhloga@student.unimelb.edu.au

Jay Acharya
1401435
acharyaj@student.unimelb.edu.au

November 2, 2023

ABSTRACT

Background: Acute Kidney Injury (AKI) is prevalent in Intensive Care Units (ICU). The ability of machine learning, using demographics, clinical comorbidities, and lab results to predict AKI outcomes, is not extensively studied.

Methods: Patients were extracted from the MIMIC-IV database based on AKI digital phenotype and grouped by mortality results. Our primary objective was to develop an ensemble machine learning model to predict AKI mortality. To achieve this, we employed feature selection by logistic regression and heatmap correlation analysis.

Results: This study involved 9,809 ICU patients, where 1,659 patients deceased. Risk factors were determined with feature selection process. The ensemble model achieved an 82.77% accuracy in predicting mortality.

Conclusions: Using machine learning with MIMIC-IV data offers a promising method to predict ICU patient mortality with AKI. This approach suggests potential for improved patient care through early prediction.

Keywords: Acute Kidney Injury, Machine Learning, MIMIC-IV, ICU Mortality.

1 Introduction

Acute Kidney Injury (AKI), defined by a rapid decline in kidney function indicated by an elevation in serum creatinine (sCr) levels or a reduction in urine output, is a significant concern associated with ICU admissions and elevated mortality rates (Srisawat et al. 2010, Alfieri et al. 2023). ICU patients diagnosed with AKI often face rapid health deterioration, emphasising the need for rigorous monitoring. However, with Electronic Health Records (EHR) data, this situation can be monitored and intervened at early stages.

While prior research has focused on AKI risk factors using various data sources and implemented machine learning to predict mortality rates in stage 3 AKI patients (Contreras-Villamizar et al. 2023, Chen et al. 2021, Liu et al. 2023, Nateghi Haredasht et al. 2023), our study introduces a novel element with the MIMIC-IV EHR dataset. With the increased data volume from MIMIC-IV dataset, this study aims to develop more robust and generalised models. Our study also conducts a thorough examination of demographic details, clinical comorbidity, and laboratory-based indicators to gain a deeper understanding of AKI risk factors.

Our goal is to predict ICU AKI patient mortality using machine learning and significant risk factors, improving patient monitoring and clinical decisions with unique data and extensive analysis.

2 Methods

The overall project pipeline is demonstrated in Figure 1. Data collection and machine learning methods were performed to achieve our final AKI mortality rate prediction model.

2.1 Data Collection and Digital Phenotyping

We employed SQL queries to retrieve EHR for the targeted AKI cohort, based on a defined set of inclusion and exclusion criteria, as well as specific demographic, comorbidity, clinical, and lab-based features from the MIMIC-IV database. This process is illustrated in Figure 1.

2.1.1 Data Resource

The study utilised data from MIMIC-IV database (Johnson et al. 2023), a retrospective database of medical records collected in the US between 2008 and 2019, comprising 431,231 admissions from 180,733 unique patients. To ethically access MIMIC-IV data, we became credentialed on PhysioNet, signed the DUA, completed human subjects training.

2.1.2 Digital Phenotype

The study operates under the assumption that patient mortality is attributable to AKI. Therefore, we have restricted our selection to patients who had only one ICU admission. Furthermore, to ensure the reliability of ICU records and that patients align with adult diagnostic benchmarks, we excluded those with brief ICU stays and individuals below 18. The table below outlines our inclusion and exclusion criteria:

Table 1: Summary of inclusion and exclusion criteria

Criteria Type	**Criteria Description**
Exclusion	Not the first ICU admission
	Length of ICU stay < 24 hours
	Patient <18 years old
	Patient with no creatinine data label
Inclusion	Increase in serum creatinine to 1.5 to 1.9 times baseline
	Increase in serum creatinine by $\geq$ 0.3 mg/dL ($\geq$ 26.5 micromol/L)
	Reduction in urine output to < 0.5 mL/kg/hour for 6 to 12 hours

Note: Due to the lack of access to patient weight data, only the first two AKI inclusion criteria were applied.

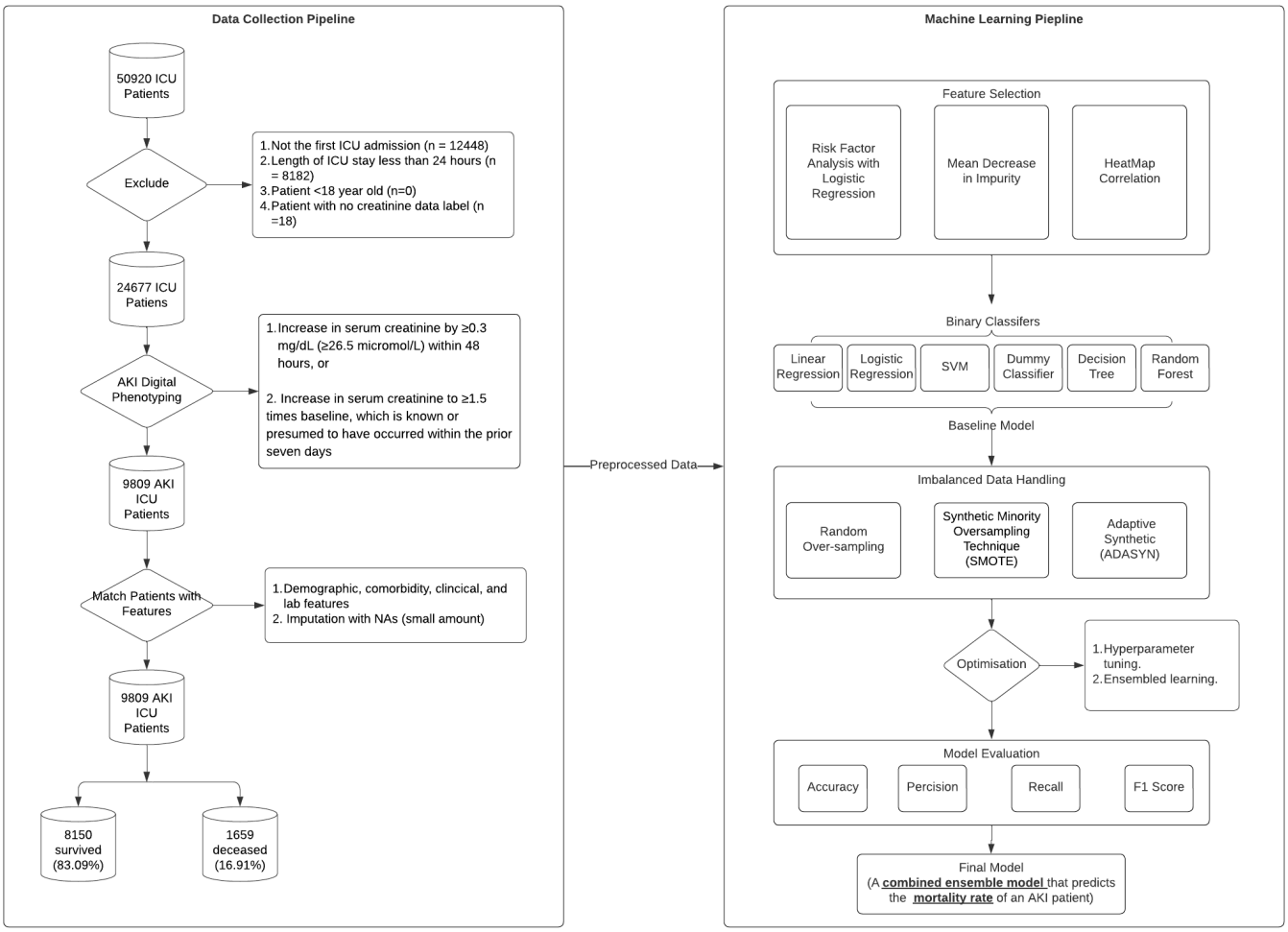


Figure 1: Project Pipeline

For early mortality risk prediction in AKI patients, we collected a comprehensive set of demographic, hematological, and biochemical measurements, including conventional and ICU-specific risk factors. Comorbidities like hypertension and diabetes were identified using ICD codes. Clinical indicators such as blood pressure, heart rate, and respiratory rate were prioritised, along with key laboratory metrics like creatinine levels, electrolyte balances, and BUN, all relevant to kidney function and AKI prediction (UpToDate 2023).

2.2 Data Preprocessing

Our data underwent a comprehensive transformation process to ensure that each row encapsulates complete information, spanning demographics, comorbidity, clinical details, and lab-based data for each individual. This involved the removal of outliers, imputation of missing values, and a rigorous feature selection step aimed at identifying the most significant predictors of mortality rates.

2.2.1 Outlier Detection

To identify and address extreme values that can adversely affect the performance of a machine learning model, we used the Interquartile Range (IQR) and box plots to visually identify and exclude outliers outside of 1.5 IQR.

2.2.2 Missing Value Imputation

To prevent potential bias or misrepresentation due to missing value in the analysis, we used the median of the feature to fill missing values. This approach assumes that the missingness is random which was confirmed by visualising the missing data patterns.

2.2.3 Feature Selection (Risk Factor Analysis)

To select the most informative features into our model, we conducted a risk factor analysis. This involves using logistic regression and point bi-serial correlation, given our binary mortality outcome. Table 2 displays the selected features and their statistical relevance to mortality. The initial selected features are based on prior research findings (Chen et al. 2021, Liu et al. 2023, Contreras-Villamizar et al. 2023).

2.3 Machine Learning

2.3.1 Baseline Model

To tackle this binary classification challenge, we employed a set of baseline models, including Linear Regression, Logistic Regression, Dummy Classifier, Decision Tree Classifier, Random Forest Classifier, and Support Vector Machine (SVM) with a linear kernel.

2.3.2 Model Optimisation

To develop a robust model and to improve upon the baseline performance, we conducted feature importance analysis and hyper-parameter tuning. Once the models are tuned, we combined them into a final ensemble classifier.

2.3.3 Imbalance Data Handling

Our dataset is significantly imbalanced, with only 16.91% of patients experiencing mortality during their ICU stay. To tackle this issue and improve model reliability, we employed three resampling techniques: Random Oversampling, Synthetic Minority Over-sampling Technique (SMOTE), and Adaptive Synthetic Sampling (ADASYN). These methods address the class imbalance problem by generating synthetic samples for the minority class, ensuring a more generalised and robust model performance.

3 Result

3.1 Data Preprocessing and Digital Phenotyping

We assembled an AKI cohort comprising 9,809 adult patients of which 8,150 survived and 1,659 passed away during their ICU stay. We exlcuded *Temperature Celsius* as more than 60% patient did not have this feature.

3.2 Risk Factor Analysis

The risk factor analysis model can be represented in the form of logistic regression as:

$$log(\frac{p(Y=1)}{p(Y=0)}) = \beta_0 + X_{demographic} \cdot \theta_1 + X_{comorbidity} \cdot \theta_2 + X_{clinic} \cdot \theta_3 + X_{lab} \cdot \theta_4$$

Where Y represents the mortality, X represents the group of features, β_0 is the intercept and θ are the coefficients to be estimated. The estimated coefficient and significance are demonstrated in Table 2.

We excluded features with no statistical significance (p>0.5): *Diabetes, Heart Rate Alarm - High,* and *SpO2 Desat Limit*. Despite studies indicating that diabetes is commonly associated with AKI, it does not appear to be a significant predictor of mortality in this analysis. This may be explained by the presence of multicollinearity among variables and potential inaccuracies in diabetes labeling, as we relied on manual filtering of diagnoses based on ICD codes. We excluded Heart Rate Alarm - High and SpO2 Desat Limit as the huge standard deviation indicate unsatisfactory data quality.

Moreover, this model aids in identifying the vulnerable group within the AKI cohort. Females are 12.91% more likely to experience the outcome compared to males, and advancing age increases susceptibility by 1.27% each year. Elevated Creatinine, INR(PT), and Respiratory Rate heighten vulnerability by 6.32%, 33.76%, and 4.45% respectively, while low Albumin levels decrease it by 44.51%. Hypertension and kidney disease mitigate vulnerability by 24.11% and 48.31% respectively. The presence of sepsis escalates it by 65.07%. These findings are pivotal for guiding clinical assessments and interventions.

Table 2: Summary of Coefficients with 95% Confidence Intervals and Significance

Category	Feature	n	%	Mean	Std Dev	Coef (Significance)	95% CI
Gender	Female (x = 0)	3972	59.50	-	-	-0.1382*	(-0.257, -0.020)
	Male (x = 1)	5837	40.49	-	-	-	-
Age	18-39	600	6.12	-	-	0.0126***	(0.008, 0.017)
	40-49	784	7.99	-	-	-	-
	50-59	1622	16.54	-	-	-	-
	60-69	2288	23.33	-	-	-	-
	70-79	2369	24.15	-	-	-	-
	80	2146	21.88	-	-	-	-
Comorbidity	Diabetes (0: No, 1: Yes)	2527	25.76	-	-	0.0734	(-0.068, 0.215)
	Hypertension (0: No, 1: Yes)	5221	53.23	-	-	-0.2759***	(-0.403, -0.149)
	Chronic Kidney Disease (0: No, 1: Yes)	3630	37.01	-	-	-0.6600***	(-0.809, -0.511)
	Sepsis (0: No, 1: Yes)	1112	11.34	-	-	0.5012***	(0.342, 0.660)
Clinical	Arterial Blood Pressure diastolic $(mmHg)$	-	-	57.02	10.25	-0.0073*	(-0.014, -0.001)
	Arterial Blood Pressure systolic $(mmHg)$	-	-	115.46	17.18	-0.0183***	(-0.022, -0.015)
	Heart Rate (bpm)	-	-	85.21	18.49	0.0159***	(0.013, 0.019)
	Heart rate Alarm - High (bpm)	-	-	135.28	1110.91	0.000	(-0.000, 0.000)
	Respiratory Rate $(insp/min)$	-	-	19.73	6.04	0.0435***	(0.034, 0.053)
	SpO2 Desat Limit $(\%)$	-	-	173.36	8636.34	-0.0008	(-0.009, 0.007)
Lab	Albumin (g/dL)	-	-	3.46	0.72	-0.5890***	(-0.671, -0.507)
	Creatinine (mg/dL)	-	-	1.79	1.78	0.0613***	(0.025, 0.097)
	Hemoglobin (g/dL)	-	-	10.27	1.75	0.0365*	(0.002, 0.071)
	INR(PT)	-	-	1.46	0.76	0.2909***	(0.215, 0.367)
	PT (sec)	-	-	16.05	8.20	0.0251***	(0.018, 0.032)
	Sodium (mEq/L)	-	-	138.39	4.95	0.0155**	(0.004, 0.027)
	Urea Nitrogen (mg/dL)	-	-	31.34	23.54	0.0108***	(0.008, 0.013)

*Note: (1) 95% confidence interval is based on the standard error; (2) *p < 0.05, **p < 0.01, ***p < 0.001.*

3.3 Machine Learning

3.3.1 Imbalance Data Handling

As illustrated in Figure 2., each resampling technique results in an equal distribution of both alive and deceased patients, achieving a balanced 1:1 ratio with slightly difference in gender balance.

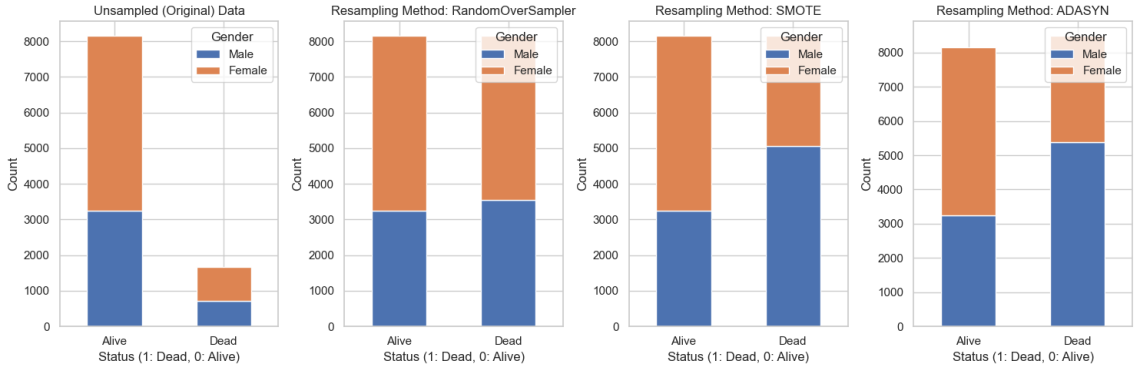


Figure 2: Imbalance Handling with Resampling Method

3.3.2 Baseline Model

In terms of baseline model performance, Linear Regression and Random Forest showed notably high accuracy, whereas SVM generated the lowest accuracy as indicated in Figure 3.

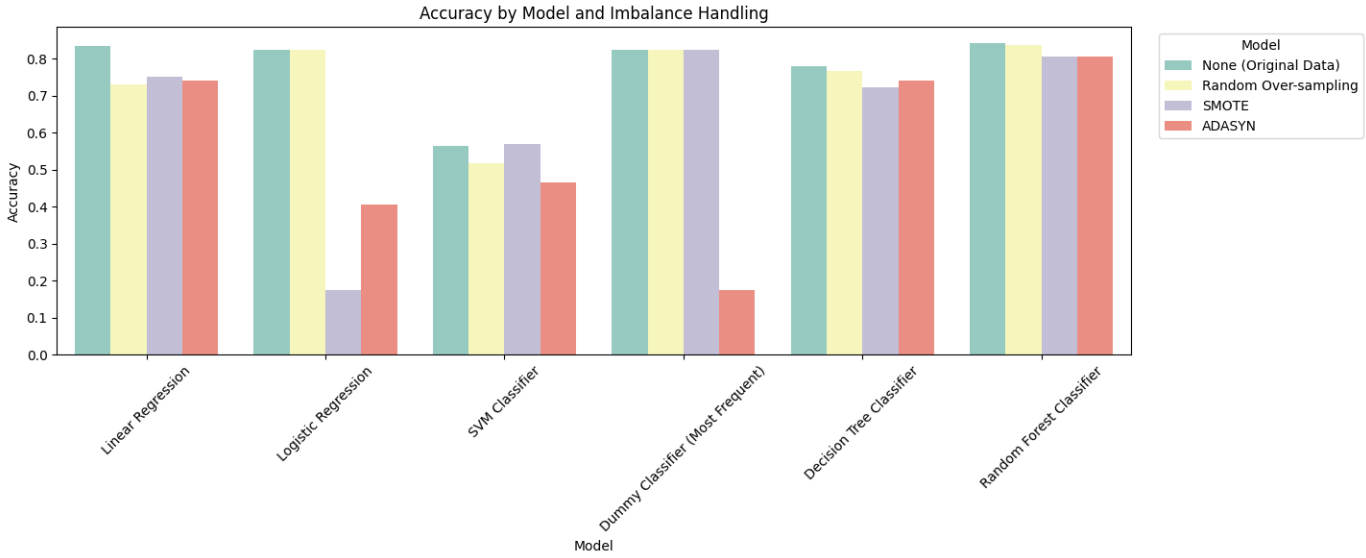


Figure 3: Baseline Model Performance

Furthermore, resampling tends to lead to reduced accuracy in general. This suggests that our model becomes more generalised as we address the potential bias associated with imbalanced labels. Overall random-resampling demonstrated best performance among unbalanced and balanced data. Additionally, the logistic regression model exhibits low accuracy with SMOTE, suggesting potential underfitting due to its simplicity. The Dummy Classifier's accuracy using ADASYN is significantly lower, indicating that the model might not capture the underlying complexities introduced into the data. Otherwise SMOTE and ADASYN have similar accuracy. In conclusion, imbalance handling techniques encompass various resampling approaches were adopted to construct more generalised model.

3.3.3 Feature Selection

As part of the model optimisation process, we have focused on ways in which features impact the model performance and how informative the selected features are during the classification process. We have conducted several experiments to gauge the importance of these features such as sequential removal to see performance differences, conducting Mean Decrease in Impurity (MDI) through the Random Forest classifier, and checking the correlation between each feature through heatmaps as shown in figure 4 and 5.

Overall we have discovered that the impact of features: 'gender', 'hypertension', 'sepsis' and 'chronic kidney disease' are negligible during the classification process. As seen in table 3, the removal of these data points seems to marginally improve the Accuracy and Precision of the model, suggesting that the influence provided by these points is similar to that of noise in the dataset. However, given the limitation in the dataset,

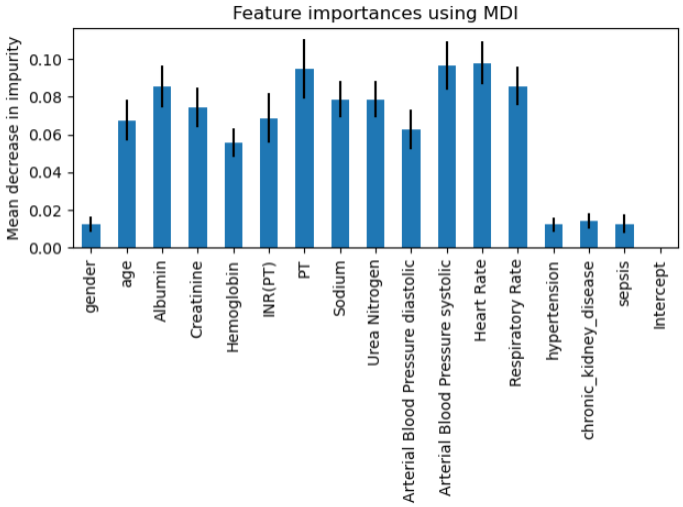


Figure 4: Feature Importance found through Mean Decrease in Impurity

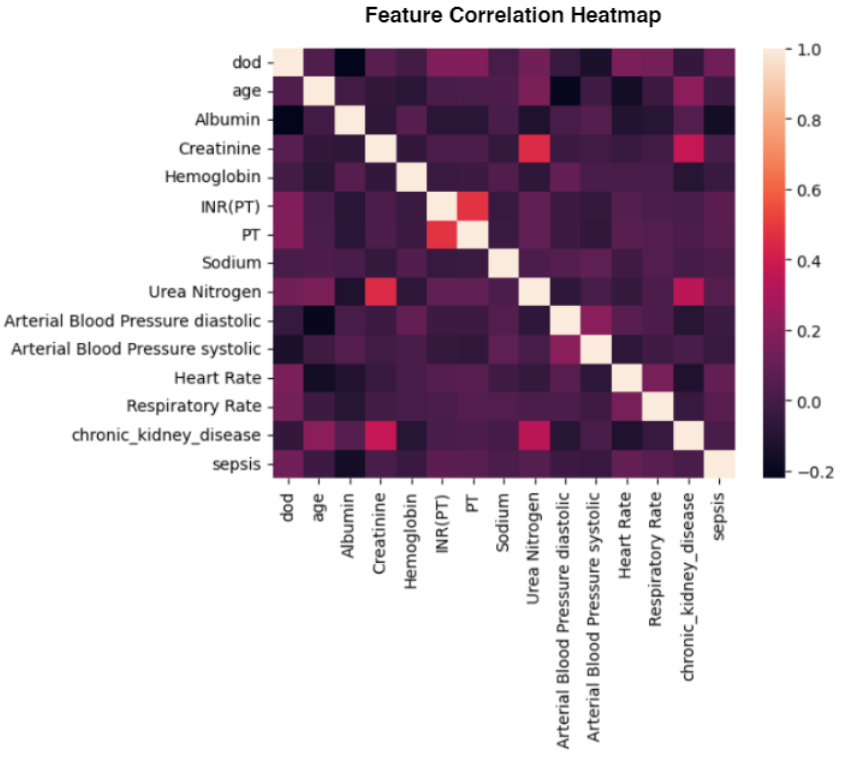


Figure 5: Correlation Heatmap of selected Features

and the imbalance between the labels, there is a possibility that we are missing informative features that correlate with the removed features, concealing a trend that may improve the model performance. Due to the negligible impact of their removal, the final models will retain these features during their training.

Table 3: Baseline LR Performance on Feature Sets

Models	Accuracy	Precision	recall	F1
All Features	0.8408	0.7277	0.0961	0.1695
Removed Features	0.8415	0.7500	0.0957	0.1694

3.3.4 Model Optimisation

Our final model, derived from an ensemble approach, was established to address the challenges presented by the imbalanced dataset discussed in the previous section. Given the imbalanced dataset explored in the previous section, adequate improvement to the classifier models was difficult to accomplish. To create a more robust model, different hyper-parameter settings and sampling methodologies were exhaustively explored for Logistic Regression, Linear Regression and Decision Tree Classifiers. Based on their accuracy, precision and recall, the top three models for each classification method were combined into a majority voting ensemble classifier. The ensemble models were further merged into a final Combined Ensemble model that balanced accuracy, precision and recall.

As seen in table 4, the accuracy did not improve, and the precision drastically fell, relative to the baseline model. This is likely due to the nature of the imbalanced dataset, where the majority voting to one label gives an average accuracy of 83% with high precision. In turn, this also resulted in incredibly low recall and f1 scores. Although the ensemble models show reduced precision, given the accuracy and recall, they represent a more robust classifier.

Table 4: Ensemble Model Performance

Ensemble Models	Accuracy	Precision	recall	F1
Baseline Linear Regression	0.8349	0.8620	0.0725	0.1337
Logistic Regression Ensemble	0.8415	0.6000	0.1898	0.2883
Linear Regression Ensemble	0.7849	0.3828	0.4428	0.4106
Decision Tree Ensemble	0.7839	0.3861	0.4699	0.4239
Combined Ensemble	0.8277	0.4874	0.3494	0.4070

4 Discussion

In the pursuit of building robust classifiers for predicting AKI patient mortality in the ICU setting, several limitations and challenges were encountered. These points will be further discussed to refine the research approach in future studies.

The missing features may potentially impact the comprehensiveness of our digital phenotype and predictive accuracy of our classifier models. Absence of weight information of parents prevents us to use features such as BMI, which can have a significant correlation with mortality. This exclusion also risks an incomplete assessment and hence for future studies obtaining comprehensive data must be prioritised to avoid missing on some critical features.

Handling sequential features effectively proved to be a challenge. In this study, we employed the starting measurement for features that had data arranged in a chronological sequence, like blood pressure over a duration. This approach may not capture the dynamic nature of patient condition efficiently. An improvement on this would be to develop new features which captures the rate of change or trend over time allowing for more insightful information, or alternatively, the use of a time-sensitive model could be considered.

We observed that removing certain features such as gender, age, hypertension, sepsis, intercept etc improved the model's performance emphasising the importance of feature engineering and selection in the medical space. In-depth evaluation is required to measure each feature's impact on classifier performance. This process should be informed by domain knowledge since it is hard for layman to make sense of the features in an informed way.

Despite our optimisation efforts, the performance of our ensemble models resulted only in marginal improvements over baseline. This finding puts a light onto the fact that the relationship between the selected features and AKI patient mortality is more complex than originally anticipated. Classifier struggled to learn the decision boundary even after refinement and feature engineering. Furthermore, The presence of an imbalanced dataset complicated model performance evaluation process as our models depend heavily on majority voting resulting in a high baseline accuracy. A more detailed evaluation technique needs to be employed to reduce any kind of bias due to this imbalance in class labels. It's important to recognise that predicting mortality for AKI involves multifaceted factors which are beyond the scope of our current research and feature set. Future studies can make use of additional data sources and incorporate better features to gain a more intricate relationship between AKI and patient mortality.

5 Conclusions

Using the MIMIC-IV database, our study experimented with machine learning models that predict mortality rates, given a dataset of ICU patients with Acute Kidney Injury. Through prepossessing, risk factor analysis, feature exploration and data sampling, we have explored different classifiers and developed an optimised ensemble model. This model provided a balanced performance, relative to the initial baseline models. However, the improvements found were marginal, with the lack of size and imbalance of the dataset hindering any potential improvements. For a future exploration of AKI mortality classifiers, we recommend the exploration of more features and predictors to create more robust and reliable models.

References

Alfieri, F., Ancona, A., Tripepi, G., Rubeis, A., Arjoldi, N., Finazzi, S., Cauda, V. & Fagugli, R. M. (2023), 'Continuous and early prediction of future moderate and severe Acute Kidney Injury in critically ill patients: Development and multi-centric, multi-national external validation of a machine-learning model', *PLOS ONE* **18**(7), e0287398.

Chen, J.-J., Kuo, G., Hung, C.-C., Lin, Y.-F., Chen, Y.-C., Wu, M.-J., Fang, J.-T., Ku, S.-C., Hwang, S.-J., Huang, Y.-T., Wu, V.-C. & Chang, C.-H. (2021), 'Risk factors and prognosis assessment for acute kidney injury: The 2020 consensus of the Taiwan AKI Task Force', *Journal of the Formosan Medical Association* **120**(7), 1424–1433.

Contreras-Villamizar, K., Barbosa, O., Muñoz, A. C., Suárez, J. S., González, C. A., Vargas, D. C., Rodríguez-Sánchez, M. P., García-Padilla, P., Valderrama-Rios, M. C. & Cortés, J. A. (2023), 'Risk factors associated with acute kidney injury in a cohort of hospitalized patients with COVID-19', *BMC Nephrology* **24**(1), 140.

Johnson, A., Bulgarelli, L., Pollard, T., Celi, L. & Mark, R. (2023), 'MIMIC-IV'.

Liu, H.-H., Wang, Y.-T., Yang, M.-H., Lin, W.-S. K. & Oyang, Y.-J. (2023), 'Exploiting Machine Learning Technologies to Study the Compound Effects of Serum Creatinine and Electrolytes on the Risk of Acute Kidney Injury in Intensive Care Units', *Diagnostics* **13**(15), 2551.

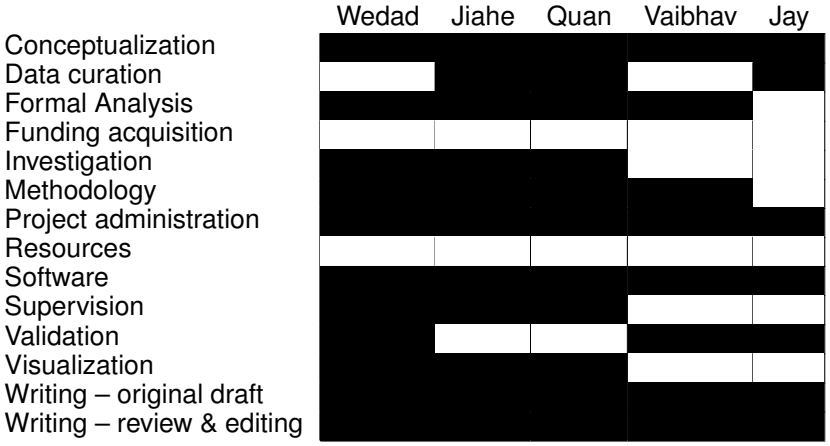
Nateghi Haredasht, F., Viaene, L., Pottel, H., De Corte, W. & Vens, C. (2023), 'Predicting outcomes of acute kidney injury in critically ill patients using machine learning', *Scientific Reports* **13**(1), 9864.

Srisawat, N., Hoste, E. E. & Kellum, J. A. (2010), 'Modern Classification of Acute Kidney Injury', *Blood Purification* **29**(3), 300–307.

UpToDate (2023), 'Kidney and patient outcomes after acute kidney injury in adults', https://www.uptodate.com/contents/kidney-and-patient-outcomes-after-acute-kidney-injury-in-adults.

6 Table of Contributions

Table 5: CRediT - Contributor Roles Taxonomy



7 GitHub Repository

For additional resources and code related to this project, please refer to our GitHub repository: `https://github.com/gracelovesyah/ml-and-health-predict-mortality-rate`