

Introduction

Acute mesenteric ischemia is a rare condition, occurring in approximately 5.3–8.4 cases per 100,000 people per year, and develops when blood flow to the small intestine becomes insufficient (Dos Santos Pereira et al., 2025). The incidence of acute mesenteric ischemia (AMI) increases with age, rising to nearly 10% in individuals older than 70 years². Delayed diagnosis of AMI is very common, since symptoms are like that of other gastrointestinal issues, but delayed treatment by more than 24 hours can lead to 60% increase in mortality (Aliosmanoglu et al., 2013). Therefore, it is crucial to detect AMI as early as possible. With rapid advances in deep learning and machine learning, AI algorithms have become sufficiently reliable for clinical application, especially in this context.

Objectives

The main objective of this study is to develop and train a deep learning logistic regression model for AMI detection, and to have this model accurate enough to be used in a clinical setting. A gap in knowledge exists in that there are no major studies showing how machine learning can be leveraged to diagnose AMI more accurately and rapidly. The central hypothesis is that the machine learning model will be able to accurately provide diagnostic results for AMI. The rationale of this study is that with further knowledge in artificial intelligence, it can start to be applied to conditions such as AMI with more confidence.

Methods

Data was taken from the MIMIC-IV dataset on BigQuery, which includes demographics and laboratory characteristics from ICU patients. The data was imported into Python and organized by sorting into AMI (n = 781) and control (n = 3128) groups, then measuring and removing missing data. The data was then analyzed by finding the median, interquartile range (IQR), counts, and percent for each variable in AMI and control. After finding the statistics, the data was cleaned up again to make a Logistic Regression model. In this model, C is 1, penalty is L2, and class weight is balanced. The model's performance was measured with AUROC, accuracy, sensitivity, specificity, positive prediction rate, negative prediction rate, and f1 score.

Results

Table 1.

Characteristic	AMI (n=781)	Controls (n=3128)	AMI missing	Control missing	p-value	p<0.05?	p<0.05? Y/N
Age	65.0 (53.0 - 77.0)	58.0 (43.0 - 71.0)	0 (0.0%)	0 (0.0%)	<0.001	Y	Y
Hematoctrit	33.8 (28.2 - 37.8)	34.4 (30.3 - 38.1)	33 (4.0%)	1004 (32.1%)	0.008	Y	Y
Hemoglobin	11.1 (8.6 - 12.1)	11.3 (8.8 - 12.8)	14 (1.7%)	1058 (33.6%)	0.101	Y	Y
White blood cell	9.3 (8.3 - 14.0)	7.8 (6.8 - 10.8)	54 (7.0%)	1087 (33.9%)	<0.001	Y	Y
Platelet count	202.0 (133.0 - 283.0)	212.0 (162.0 - 271.0)	54 (7.0%)	1042 (33.0%)	0.030	Y	Y
Creatinine	1.0 (0.7 - 1.4)	0.9 (0.7 - 1.3)	55 (7.0%)	1069 (34.0%)	0.021	Y	Y
Blood urea nitrogen	20.0 (12.0 - 33.0)	16.0 (12.0 - 24.0)	59 (7.0%)	1081 (34.0%)	<0.001	Y	Y
Sodium	138.0 (130.0 - 145.0)	138.0 (137.0 - 143.0)	53 (6.0%)	1084 (34.7%)	0.003	Y	Y
Potassium	4.0 (3.8 - 4.4)	4.1 (3.7 - 4.4)	51 (6.0%)	1091 (33.9%)	0.008	Y	Y
Cholesterol	166.0 (100.0 - 198.0)	169.0 (100.0 - 198.0)	53 (6.0%)	1087 (34.0%)	<0.001	Y	Y
Bicarbonate	24.0 (21.0 - 28.0)	23.0 (22.0 - 27.0)	40 (5.0%)	1110 (35.0%)	<0.001	Y	Y
Arden gap	14.0 (12.0 - 17.0)	13.0 (11.0 - 16.0)	60 (7.0%)	1111 (35.0%)	<0.001	Y	Y
Glucose	122.0 (90.0 - 180.0)	108.0 (80.0 - 138.0)	43 (5.0%)	1128 (36.0%)	<0.001	Y	Y
Sex					0.227	N	N
F	389 (28.0%)	1618 (51.0%)					
M	392 (30.2%)	1482 (47.1%)					
Race					<0.001	Y	Y
White	579 (75.0%)	2105 (67.0%)					
Black	73 (9.3%)	624 (19.7%)					
Asian	28 (3.0%)	68 (3.1%)					
Hispanic/Latino	18 (2.3%)	121 (4.2%)					
Other/Unknown	86 (11.1%)	288 (9.2%)					
Admission type					<0.001	Y	Y
Emergency/ICU	682 (87.3%)	1455 (46.5%)					
Observation	51 (6.5%)	1519 (48.1%)					
Elective/Surgical	48 (6.1%)	283 (9.0%)					
Other	0 (0.0%)	0 (0.0%)					

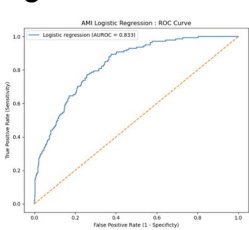
In the table, the median and IQR (25% - 75%) were recorded for numeric variables and the counts and percentages of and . p-values were assessed using the Mann–Whitney U test for continuous measures and the Chi-square test for categorical measures, with missing values recorded for each group. Nearly all variables met the conventional p < 0.05 threshold. Missing data were more common in the control group, likely reflecting the urgency of AMI care, where clinicians order more tests to support rapid diagnosis and treatment.

Table 2.

	Metric	Value
0	AUROC	0.833
1	Accuracy	0.763
2	Balanced accuracy	0.742
3	Sensitivity	0.709
4	Specificity	0.775
5	PPV	0.412
6	NPV	0.923
7	F1 score	0.521

The model outputs several performance metrics: AUROC, Accuracy, Balanced Accuracy, Sensitivity, Specificity, PPV, NPV, and F1 Score. The ROC curve plots the True Positive Rate against the False Positive Rate at different thresholds (see Figure 1). Accuracy is the proportion of correct predictions. Sensitivity measures correct predictions among AMI patients. Specificity measures correct predictions among control patients. Balanced Accuracy is the average of Sensitivity and Specificity. PPV is the proportion of correct positive predictions. NPV is the proportion of correct negative predictions. The F1 Score is the harmonic mean of PPV and Sensitivity.

Figure 1.



This the Receiver Operating Characteristics Curve. It plots all possible threshold values by the True Positive rate (Y-Axis) and the False Positive Rate (X-Axis). The true positive rate is the number of cases with AMI that the model gets right out of all cases with AMI. The False Positive Rate is the number of cases with the disease the model gets wrong out of all cases with the condition. The best possible value is in the top left of the graph with 0 positive cases gotten wrong and 100% of the positive cases gotten right. The area under the graph can be used to determine the accuracy because the more the curve “bulges” towards (0,1), the larger the AUC will be.

Discussion

In Table 1, several meaningful differences were observed. AMI patients were older, with a median age of 65.0 years compared with 58.0 years in controls. Laboratory values also differed significantly: AMI patients showed higher white blood cell counts, blood urea nitrogen levels, and glucose levels. Most AMI admissions were in the Emergency/Urgency category with 682 (87.3%) versus 51 (6.5%) in Observation and 48 (6.1%) in Elective/Surgical. Sex distribution was similar between groups, indicating no significant association. In the results table (Table 2), the AUROC is 0.8, which is better than my hypothesis, but the specificity and sensitivity are a little lower than expected at with the average being around 0.74. Having a high sensitivity is especially important because getting a false negative can be more fatal than getting a false positive when detecting diseases. The PPV is also lower than expected. Overall, these findings show a mediocre to good score for my model.

Conclusion

These results show how AI and ML could be used in many fields with enough training and push for them. To restate; my hypothesis was that the model would be able to accurately predict whether a person has AMI or not up to 0.75 AUC. I don't think this hypothesis was proven true, because even though it surpassed my expectation on the AUROC, everything else is mediocre. Future studies in this area could try to use different models like a Decision Tree or Random Forest. Some limitations of my project are that it has a relatively small sample size because of AMI's rarity and my oversampling.