INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 16 No. 2 (2024) 114-122
https://publisher.uthm.edu.my/ojs/index.php/ijie

Machine Learning Classifications of Multiple Organ Failures in a Malaysian Intensive Care Unit

Norliyana Nor Hisham Shah[1], Normy Norfiza Abdul Razak[1]*, Athirah Abdul Razak[1], Asma' Abu-Samah[2], Fatanah M. Suhaimi[3], Ummu Jamaluddin[4]

[1] *Institute of Energy Infrastructure (IEI), College of Engineering, Universiti Tenaga Nasional, Kajang, Selangor, 43000, MALAYSIA*
[2] *Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor, 43600, MALAYSIA*
[3] *Advanced Medical and Dental Institute, Universiti Sains Malaysia, Kepala Batas, Penang, 13200, MALAYSIA*
[4] *College of Engineering, Universiti Malaysia Pahang, Pekan, Pahang, 26000, MALAYSIA*

*Corresponding Author: Normy@uniten.edu.my
DOI: https://doi.org/10.30880/ijie.2024.16.02.012

Article Info

Received: 19 November 2023
Accepted: 2 January 2024
Available online: 29 April 2024

Keywords

Multiple organ failures, machine learning, classifications, intensive care unit

Abstract

Multiple organ failures are the main cause of mortality and morbidity in the intensive care unit (ICU). The progression of organ failures in the ICU is usually monitored using the Sequential Organ Failure Assessment (SOFA) score. This study aims to perform the classification of multiple organ failures using machine learning algorithms based on SOFA score. Ninety-eight ICU patients' data were obtained retrospectively from Universiti Malaya Medical Centre for analysis. Several machine learning algorithms which are decision tree, linear discriminant, naïve Bayes, support vector machines, k-nearest neighbor, AdaBoost, and random forest were used for the classification. The classifiers were trained on 80% of the patients with 10-fold cross-validations and assessed on 20% of patients using 34 variables in the ICU. The random forest algorithm was able to achieve 99.8% accuracy and 99.9% sensitivity in the training dataset. Meanwhile, the AdaBoost algorithm achieved 99.1% sensitivity in the testing dataset. This study demonstrates the performances of different machine learning algorithms in the classification of multiple organ failures. The feature selection shows respiratory rate and mean arterial pressure (MAP) as the most important variables using chi-square test while insulin and fraction of oxygenated hemoglobin are the most important predictors by the mutual information test.

1. Introduction

Multiple organ failures (MOF) are defined as the presence of two or more organ dysfunctions simultaneously. The term is sometimes interchangeably used with multiple organ dysfunction syndrome to describe improving organ function after receiving treatment. A more known representation of organ failure in the ICU are such as acute respiratory distress syndrome (ARDS), disseminated intravascular coagulation (DIC), or acute kidney injury (AKI). Sepsis is the main cause of organ failure as a response to infection, and septic shock is usually described for patients with multiple organ failures [1]. Other causes such as burn, trauma, and hematologic malignancies

patients also developed multiple organ failures [2-5]. In critically ill patients with COVID-19, MOFs are the main cause of mortality [6].

Several risk scores were developed to assess organ failure and mortality among patients in the ICU. The most used score for organ failure is the Sequential Organ Failure Assessment (SOFA) score [7]. Another known severity score for organ failure is the multiple organ dysfunction score (MODS) [8]. Both these scores evaluate the same organ systems which are respiratory, cardiovascular, renal, hepatic, coagulation, and central nervous system. These severity scores are preferred as they use a single variable to monitor each organ failure progression. In SOFA score, each organ is given a score between 0 to 4, where a score of 4 indicates severe organ dysfunction.

The SOFA score is monitored daily using the worst reading of variables associated with each organ. Unlike the Acute Physiologic Assessment and Chronic Health Evaluation II (APACHE II) or Simplified Acute Physiology Score II (SAPS II) which predict the risk of mortality based on variables value at admission, the daily SOFA score is accumulated for the six organs out of 24 total scores [9,10]. The individual score is used to monitor organ failure progression, while the cumulative daily SOFA score was mainly used as predictor of mortality. The cumulative SOFA score on admission up to 11 predicted risks of mortality at 50% while patients with SOFA score more than 11 on admission have a risk of mortality up to 95% [11,12]. However, there is no clear indication of the number of organ failures using the cumulative SOFA score. For example, a daily SOFA score of 12 can be cumulative of 3 organ systems with a score of 4 for each organ. The cardiovascular and renal components of the SOFA score are the highest predictors for 1-year mortality when evaluated independently [13].

The surge of electronic health records data has spurred machine learning applications in healthcare. In MOF prediction, machine learning has been utilized to predict the incidences within trauma and sickle cell disease patients [14,15]. Otherwise, machine learning algorithms have been used to predict mortality in patients with MOF [16]. Meanwhile, early sepsis prediction has shown outstanding potential with ground-breaking outcomes using machine learning [17-19]. A study has shown that sepsis can be predicted up to 48 hours in advance with high accuracy and minimal predictors [20]. These studies collectively highlight the potential of machine learning in identifying and predicting risk factors associated with specific organ failures in critically ill patients.

In this study, we aim to report the classification of multiple organ failures based on SOFA score using machine learning. The main contributions of this paper include:

- This article reports the performances of machine learning algorithms in the classification of multiple organ failures among ICU patients in Malaysia.
- Patients' comorbidities and metabolic variations were included as variables in the classification

2. Materials and Methods

2.1 Patients' Data and Pre-Processing

Patients' data were collected retrospectively from Universiti Malaya Medical Centre (UMMC) for patients staying in the intensive care unit (ICU) from February to October 2018. Ninety-eight patients from age 17 to 82 years old with an average age of 57.14 consisting of 61 male patients were included in this study. Demographics data included are age and gender of the patients, meanwhile, the comorbidities considered as variables for classification are diabetes mellitus and hypertension. However, other comorbidities of the patients such as ischemic heart disease, coronary artery disease, ventricular failure, atrial fibrillation, congestive cardiac failure, chronic obstructive pulmonary disorder, obstructive sleep apnea, interstitial lung disease, bronchial asthma, chronic kidney disease, and end-stage renal failure were also recorded. Vital signs were recorded for all patients. Other blood gas measurements including pH, arterial blood gas, electrolytes, oximetry values, acid-base status, and metabolic variations variables were extracted whenever available. The total number of days for all patients was 628 days (13241 hours available). Table 1 shows the details of 34 variables considered for the machine learning algorithms.

Prior to feeding the data for machine learning classifications, several data pre-processing was performed. The data was cleaned and stratified before being trained and tested. The patients' gender was transformed into binary where male is 1, and female (2). Meanwhile, the presence of diabetes mellitus (DM) and hypertension (HPT) was recorded as 1 and the absence as 0. Similarly, for patients under mechanical ventilation and insulin infusion was labelled as 1. Other types of variables such as vital signs, blood gas values, and other variables were available hourly. The missing values for vital signs and glucose readings were imputed as carry forward until a new measurement is available while the other variables were recorded when available. The PF ratio was imputed when both partial pressure of oxygen (PO2) and fraction of inspired oxygen (FiO2) readings were available in the same hour. The snippet of variables used for classifications is shown in Table 2. The table included number of rows as row identifier and the outcome labelled as 0 for non-failure and 1 for multiple organ failures (MOF).

Int. Journal of Integrated Engineering Vol. 16 No. 2 (2024) p. 114-122

2.2 Classifications

The primary outcome for classification is multiple organ failures which is defined as 2 or more organ failures based on SOFA score. An organ system is considered as failure for SOFA score ≥2. Table 2 shows the metrics and variables for organ failure measurements in SOFA score. The outcome is labelled as 1 for patients with multiple organ failures denoted as MOF. Meanwhile, patients without organ failures or with single organ failure are labelled as 0 denoted as NF in the confusion matrix.

Table 1 *Variables information*

Type of features	No.	Details
Demographic	2	Age, gender
Comorbidities	2	Diabetes mellitus (DM), hypertension (HPT)
Vital signs	7	Temperature (temp), heart rate (HR), respiratory rate (RR), systolic blood pressure (sys), diastolic blood pressure (dia), mean arterial pressure (MAP), arterial oxygen saturation (SpO2)
Blood gas values	17	pH, partial pressure of oxygen (PO2), partial pressure of carbon dioxide (PCO2), Electrolytes: Potassium (K), sodium, calcium chloride (Cl), Oximetry values: total hemoglobin (Hb), fractional oxygen saturation (fraction of oxygenated hemoglobin, sO2), and its hemoglobin fractions: FO2Hb, FCOHb, FMetHb, Acid base status: Base excess (BE), bicarbonate (HCO3), haemotocrit, glucose, and lactace.
Others	6	Use of mechanical ventilation, provided insulin, fraction of inspired oxygen (FiO2), urine output, PO2/FiO2 (PFratio), and insulin sensitivity (SI).

Table 2 *Sample rows of data included for classification*

No	Age	Gender	DM	HPT	SI	glucose	insulin (Y/N)	HR	Vent (Y/N)	RR	sy	di
1	67	1	1	1	0.000255396	8	0	60	0	24	87	55
2	67	1	1	1	0.000255396	8	0	65	0	24	92	51
3	67	1	1	1	0.000255396	8	0	65	0	23	101	61
4	67	1	1	1	0.000255396	8	0	64	0	23	111	58
5	67	1	1	1	0.000255396	8	0	64	0	23	111	58
6	67	1	1	1	0.000255396	9.8	0	64	0	23	111	58
7	67	1	1	1	8.47E-06	10.1	0	78	1	15	145	58
8	67	1	1	1	8.47E-06	10.1	0	53	1	21	121	53
9	67	1	1	1	2.74E-05	10.1	0	55	1	16	129	59
10	67	1	1	1	4.85E-05	14.6	0	51	1	16	108	49
11	67	1	1	1	5.15E-06	14.6	1	64	1	17	122	54
12	67	1	1	1	5.92E-06	14.3	1	61	1	16	108	46
13	67	1	1	1	6.63E-06	14.3	1	60	1	16	94	46

Table 3 *SOFA Score (adapted from [7])*

Score	0	1	2	3	4
Respiratory					
PO2/FiO2, mmHg (kPa)	≥400 (53.3)	<400 (53.3)	<300 (40)	<200 (26.7)*	<100 (13.3)*
Coagulation					

Platelets x 103/µL	≥150	100 – 149	50 – 99	20 – 49	<20
Hepatic					
Bilirubin, µmol/L	<20	20 – 32	33 – 101	102 – 204	>204
Cardiovascular	MAP ≥ 70 mmHg	MAP < 70 mmHg	Low dose dopamine or any dose dobutamine	Low-medium dose noradrenaline or adrenaline, medium dose dopamine	High dose noradrenaline, adrenaline, or dopamine
Renal					
Creatinine, µmol/L	<110	110 – 170	171 – 299	300 – 440	>440
Urine output	-	-	-	<500 mL/day	<200 mL/day

*With respiratory support; PO_2 = partial pressure of oxygen, FiO_2 = fraction of inspired oxygen.

Several machine learning algorithms were considered for classification. They are decision tree (DT), linear discriminant (LD), naïve Bayes (NB), support vector machine (SVM), k-nearest neighbor (KNN), with two ensemble algorithms which are AdaBoost and random forest (RF). Decision tree (DT) classifier starts from a root node at the base of the tree which splits into branches. Each branch is attached to a leaf node representing the predicted outcome. In order to reach the final outcome, follow the decision from the root node to the desired leaf node. The usage of decision tree is preferable as they are easy to interpret and require lower memory usage. However, they are prone to overfitting. Linear Discriminant (LD) works by finding the combination of features with linear relationships to separate between the classes. The features in the linear discriminant analysis were assumed to be normally distributed. Naive Bayes (NB) classifier uses the Bayes theorem to calculate the probability of each class given the probability of other events. The advantage of naive Bayes classifier is they can handle a high number of features as each feature is considered as conditionally independent of each other. Support vector machines (SVM) separates data points into two regions using the best hyperplane, which has the highest margin between the closest data point and the hyperplane. In k-nearest neighbour (kNN), data is classified based on the majority class of its neighbours, where k is the number of neighbours considered. A weak classifier uses a higher number of neighbours, while a lower neigbours signifies a fine classifier. The ensemble learning method, AdaBoost assigns weight to misclassified samples and then, iteratively train weak classifiers focusing more on these misclassified samples. The sum of these weighted weak learners forms a final stronger classifier. Meanwhile, the random forest (RF) is a preferred classifier as they reduce overfitting by training random subset of features on one tree and another subset on another tree. The average prediction of these multiple decision makes the ensemble final prediction.

The performances of these algorithms were measured using accuracy, sensitivity, specificity, and precision. These performance measures were calculated based on elements in the confusion matrix. The confusion matrix for two-class classifications and the performance measures are shown in Table 4. In the confusion matrix, the rows refer to actual number of cases while the columns are predicted number of cases. True Positive (TP) refers to correctly identified cases of multiple organ failures, while True Negative (TN) are non-failure cases identified correctly by the algorithms. Meanwhile, False Positive (FP) refers to patients without MOF which are mistakenly identified as experiencing multiple organ failures. Conversely, False Negative (FN) refers to incorrect classifications of patients with MOF. The dataset is divided randomly into 80% training data and 20% testing data. The classification algorithms are subjected to 10-fold cross-validations in the training dataset before running the model for the testing dataset.

Table 4 *Confusion matrix and performance measures*

		Predicted Class	
		MOF	NF
Actual Class	MOF	True Positive (TP)	False Negative (FN)
	NF	False Positive (FP)	True Negative (TN)

Accuracy	ACC = (TP + TN) / (TP + TN + FP +FN)	Correctly identified over all cases
Sensitivity	TPR = TP / (TP +FN)	Correctly identified as positive (actual)
Specificity	SPE = TN / (TN + FP)	Correctly identified negative

Int. Journal of Integrated Engineering Vol. 16 No. 2 (2024) p. 114-122

Precision	PPV = TP / (TP + FP)	Correctly identified positive (prediction)

3. Results and Discussion

3.1 Structure

The demographics of 98 patients considered in this study are shown in Table 5. Patients with multiple organ failures (MOF) are generally older with median age of 62 as compared to 54 years old for patients without MOF, and 59 for the whole patient cohort. There are 62 male patients and 68 patients experienced multiple organ failures during their stay in the ICU. Distribution among race in the whole patient cohorts are balanced among Malay, Chinese, and Indian. Three other patients were identified as non-Malaysians. Among the comorbidities, diabetes mellitus and hypertension were prominent among all patients at 59% and 52%, respectively. Among all patients, 38 have both diabetes and hypertension. In the MOF cohort, there are higher number of patients with comorbidities compared to patients without MOF. For example, MOF patients with diabetes consist of 76% of patients with diabetes as comorbidity. Similarly, 31% of MOF patients have both diabetes and hypertension in the whole patient cohort. It is worth noting that 15% of patients have no reported comorbidities and pulmonary-related comorbidities have the lowest number of patients at 7%.

The breakdown of organ failure experienced by the patients during their stay in the ICU is shown in Fig. 1. This figure exhibited the number of patients with SOFA score ≥2 for each organ shown. Most patients experienced respiratory organ failure (PFratio<300) followed by cardiovascular failure. Patients with renal failure (39%) were determined by level of creatinine more than 170 μmol/L while bilirubin level >32 μmol/L were considered as hepatic failure. Subsequently, MOF patients were identified as patients with 2 or more organ failures.

Table 5 *Patients' demographics*

Demographics	All (n = 98)		NF (n = 30)	MOF (n = 68)
Age (median)	17 – 82	(59)	17 – 80 (54)	21 – 82 (62)
Gender				
• Male	62		20	42
• Female	36		10	26
Race				
• Malay	35	(35.71%)	10	25
• Chinese	36	(36.73%)	12	24
• Indian	24	(24.49%)	6	18
• Others	3	(3.07%)	2	1
Comorbidity*				
• Diabetes Mellitus (DM)	58	(59.18%)	14	44
• Hypertension (HPT)	51	(52.04%)	14	37
• Cardiovascular	21	(21.43%)	6	15
• Pulmonary	7	(7.14%)	2	5
• Renal	13	(13.27%)	1	12
• No listed comorbidity	15	(15.31%)	9	6
• DM & HPT	38	(38.78%)	8	30

*Cardiovascular comorbidity includes ischemic heart disease, coronary artery disease, ventricular failure, atrial fibrillation, congestive cardiac failure, pulmonary comorbidity includes obstructive sleep apnea, interstitial lung disease, bronchial asthma, and chronic obstructive pulmonary disorder (COPD), renal comorbidity includes chronic kidney disease (CKD) or end-stage renal failure (ESRF).

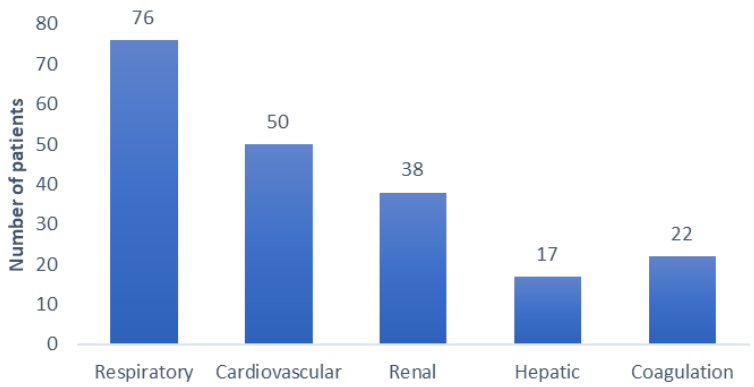

Fig. 1 *Number of patients with organ failure*

The training and testing dataset used were the same for each classification algorithm. Fig. 2 shows the performances and confusion matrices as results of classifications. The highest accuracy in the whole and training dataset were achieved using random forest classifier at 99.8%. The average accuracy for the other classifiers is 81.12% using the whole dataset and 85.4% for the training dataset. Accuracy measures correctly identified both positive and negative cases among all number of cases. Out of 100 cases, 80% accuracy shows that 80 cases were predicted accordingly as positive or negative. The random forest classifier also achieved the highest sensitivity, specificity, and precision in the whole and training dataset. At nearly 100% for these three performance matrices show that the classifier was able to classify all cases accordingly.

The average sensitivity for all classifiers using whole, training, and testing datasets is 97.8%. High sensitivity is attributed to the ability of the classifier to identify correct MOF incidence over the actual number of MOF cases. In this case, it suggests that the predictor variables selected for the classifiers were able to classify MOF incidences correctly. The average precision for the whole and training dataset for all classifiers are 83.75% and 87.63%, respectively. However, the average precision across all classifiers in the testing dataset is 50.62%. The average precision in the testing dataset is significantly reduced compared to the whole and training dataset. Precision measures the ability of the classifier algorithm to correctly classify positive cases as compared to the total positive classification (actual and positively predicted). This contrasts with sensitivity which measures positive classification based on actual number of positive cases. In the testing dataset, a whole number of non-MOF cases were identified as false positives increasing the number of positives classification. Thus, the ratio between the correctly predicted and total positive prediction is much lower than in the training dataset. Meanwhile, in the whole and training dataset, the number of MOF cases correctly predicted is higher (TP) than false positives, increasing the precision matrix. This also explains the lower specificity achieved by all the classifiers using whole, training, and testing dataset. As specificity measures the correctly identified negative cases (in this case, patients without MOF), the classifiers achieved low specificity as they are prone to classify the negative cases as positive. It is not unusual for a model to have high sensitivity and low specificity as their relationship is inversely related. Higher sensitivity is usually preferred when the model is intended for positive classification (patients with MOF). Thus, these results suggest that the predictor variables selected were suitable for identifying MOF incidences. However, lower specificity model suggests overestimated patients without MOF as having MOF and may create false alarms.

Int. Journal of Integrated Engineering Vol. 16 No. 2 (2024) p. 114-122

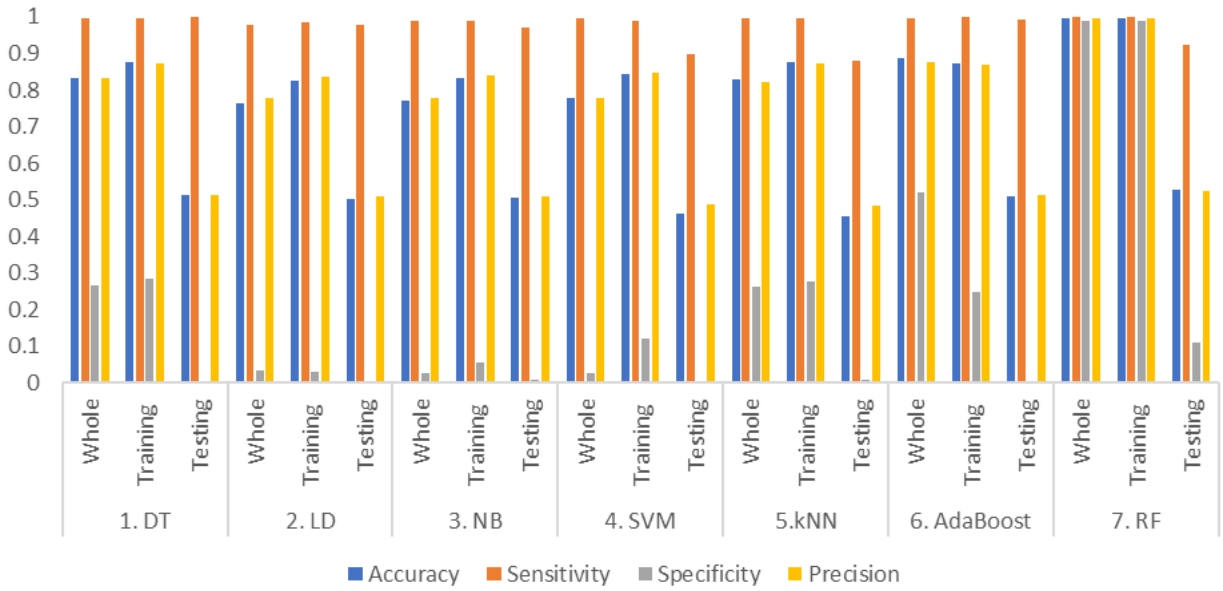

Fig. 2 *Performances of machine learning algorithms*

The most relevant variables as predictors were investigated using the chi-square and mutual information methods. The most important predictors using chi-square test are respiratory rate and mean arterial pressure (MAP) which are not included in Fig. 3a) as they are given infinite scores. The next most important feature is age, followed by temperature, systolic, and diastolic blood pressure. Diabetes status and insulin sensitivity as metabolic variation variables were also included among 10 highest predictor ranking using chi-square test. Meanwhile, the mutual information feature selection method shows age as the most important variable. The use of ventilation and PF ratio are also ranked as the most important variables. The overlapping variables as identified using the chi-square test and mutual information are age, temperature, systolic, and diastolic blood pressure, and diabetes status. Fig. 3 shows the scores of the variable's importance using chi-square test and feature information. The limitation in this study is small sample obtained from a single center adult ICU in Malaysia. This might hinder generalizability of the machine learning models to other types of care such as multi-centre ICUs, emergency, inpatient, or pediatrics. There is also no status of mortality among the patients thus, limiting the association between organ failures and mortality. Lastly, this study only includes bedside monitoring variables omitting laboratory values which might be important biomarkers for MOF. Therefore, it is suggested to include other variables including laboratory values as predictors and perform feature selection towards reliable and optimal performance models.

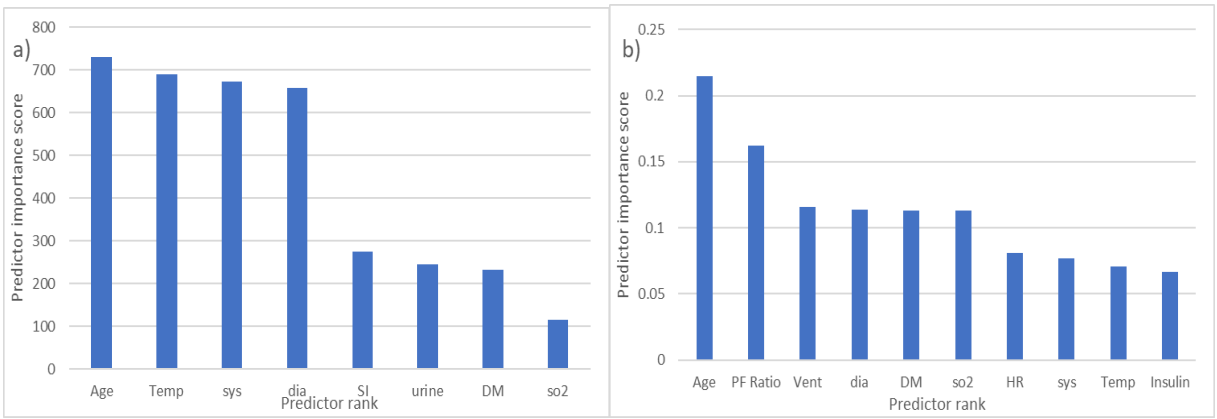

Fig. 3 *Feature selection using (a) Chi-square test; (b) Mutual information*

Table 6 *Confusion matrices*

		Whole		Training		Testing	
1. Decision tree		MOF	NF	MOF	NF	MOF	NF
	MOF	10220	39	8995	26	1237	1
	NF	2189	792	1295	520	1167	0
2. Linear discriminant		MOF	NF	MOF	NF	MOF	NF
	MOF	10031	228	8885	136	1210	28
	NF	2879	102	1761	54	1166	1
3. Naïve Bayes		MOF	NF	MOF	NF	MOF	NF
	MOF	10142	117	8905	116	1203	35
	NF	2896	85	1714	101	1157	10
4. SVM		MOF	NF	MOF	NF	MOF	NF
	MOF	10236	23	8914	107	1110	128
	NF	2898	83	1592	223	1165	2
5. kNN		MOF	NF	MOF	NF	MOF	NF
	MOF	10214	45	8994	27	1090	148
	NF	2197	784	1314	501	1158	9
6. AdaBoost		MOF	NF	MOF	NF	MOF	NF
	MOF	10209	50	9011	10	1227	11
	NF	1432	1549	1366	449	1167	0
7. Random forest		MOF	NF	MOF	NF	MOF	NF
	MOF	10252	7	9017	4	1143	95
	NF	35	2946	22	1793	1036	131

4. Conclusion

This article discussed different machine learning algorithms as classifiers for patients with MOF. The results show that the random forest classifier was able to correctly identify MOF cases with 99% accuracy in the whole and training dataset. The machine learning classifiers using the variables were able to achieve 98% sensitivity. The high sensitivity suggests the ability of the predictor variables to correctly identify all MOF incidences.

Acknowledgement

We would like to thank Ministry of Higher Education for the Fundamental Research Grant Scheme (FRGS/1/2019/STG05/UNITEN/02/1). We acknowledge Universiti Malaya Medical Centre for providing the patients data used for analysis.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *NNHS, NNAR;* ***data collection:*** *NNHS, AAR, AAS;* ***analysis and interpretation of results:*** *NNHS;* ***draft manuscript preparation:*** *NNHS, NNAR, AAR, AAS, FMS, UJ. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Seymour, C. W., Liu, V. X., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., ... & Angus, D. C. (2016). Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA, 315(8), 762-774.

[2] Mannes, M., Schmidt, C. Q., Nilsson, B., Ekdahl, K. N., & Huber-Lang, M. (2021). Complement as driver of systemic inflammation and organ failure in trauma, burn, and sepsis. Seminars in Immunopathology (pp. 1-16). Springer Berlin Heidelberg.

[3] Zhang, P., Zou, B., Liou, Y. C., & Huang, C. (2021). The pathogenesis and diagnosis of sepsis post burn injury. Burns & Trauma, 9.

[4] de Vries, V. A., Mueller, M. C., Arbous, M. S., Biemond, B. J., Blijlevens, N. M., Kusadasi, N., ... & van den Bergh, W. M. (2019). Long-term outcome of patients with a hematologic malignancy and multiple organ failure admitted at the intensive care. Critical Care Medicine, 47(2), e120.

[5] Cole, E., Gillespie, S., Vulliamy, P., Brohi, K., Akkad, H., Apostolidou, K., ... & Welters, I. (2020). Multiple organ dysfunction after trauma. *Journal of British Surgery*, 107(4), 402-412.

[6] Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., ... & Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet, 395(10223), 507-513.

[7] Vincent, J. L., Moreno, R., Takala, J., Willatts, S., De Mendonça, A., Bruining, H., ... & Thijs, L. G. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure: On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Medicine, 22, 707-710.

[8] Marshall, J. C. (1995). Multiple organ dysfunction syndrome (MODS). Clinical trials for the treatment of sepsis, 122-138.

[9] Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: a severity of disease classification system. Critical Care Medicine, 13(10), 818-829.

[10] Le Gall, J. R., Lemeshow, S., & Saulnier, F. (1993). A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA, 270(24), 2957-2963.

[11] Ferreira, F. L., Bota, D. P., Bross, A., Mélot, C., & Vincent, J. L. (2001). Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA, 286(14), 1754-1758.

[12] Do, S. N., Dao, C. X., Nguyen, T. A., Nguyen, M. H., Pham, D. T., Nguyen, N. T., ... & Luong, C. Q. (2023). Sequential Organ Failure Assessment (SOFA) Score for predicting mortality in patients with sepsis in Vietnamese intensive care units: a multicentre, cross-sectional study. BMJ open, 13(3), e064870.

[13] Ralib, A. M., Said, N. A., Ramly, N. F., Nanyan, S., Ismail, M. N., & Nor, M. B. M. (2022). Serial Evaluation of Sequential Organ Failure Assessment Score in Predicting 1-Year Mortality in Critically Ill Patients. IIUM Medical Journal Malaysia, 21(3).

[14] Wang, Y., Zhao, Y., Callcut, R., & Petzold, L. (2021). Empirical analysis of machine learning configurations for prediction of multiple organ failure in trauma patients. arXiv preprint arXiv:2103.10929.

[15] Mohammed, A., Podila, P. S., Davis, R. L., Ataga, K. I., Hankins, J. S., & Kamaleswaran, R. (2020). Using machine learning to predict early onset acute organ failure in critically ill intensive care unit patients with sickle cell disease: retrospective study. Journal of Medical Internet Research, 22(5), e14693.

[16] Liu, X., Hu, P., Mao, Z., Kuo, P. C., Li, P., Liu, C., ... & Zhou, F. (2020). Interpretable machine learning model for early prediction of mortality in elderly patients with multiple organ dysfunction syndrome (MODS): a multicenter retrospective study and cross validation. arXiv preprint arXiv:2001.10977.

[17] Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., ... & Das, R. (2016). Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Medical Informatics, 4(3), e5909.

[18] Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. Critical Care Medicine, 46(4), 547.

[19] Giannini, H. M., Ginestra, J. C., Chivers, C., Draugelis, M., Hanish, A., Schweickert, W. D., ... & Umscheid, C. A. (2019). A machine learning algorithm to predict severe sepsis and septic shock: development, implementation and impact on clinical practice. Critical Care Medicine, 47(11), 1485.

[20] Barton, C., Chettipally, U., Zhou, Y., Jiang, Z., Lynn-Palevsky, A., Le, S., ... & Das, R. (2019). Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. Computers in Biology and Medicine, 109, 79-84.