

An Accurate Data Preparation Approach for the Prediction of Mortality in ACLF Patients using the CANONIC Dataset

M.S. Garcia*, Gordon Doyle*, Gabriele Ranco*, Vicente Arroyo†, Marco Pavesi†, Elizabet Garcia†, Faouzi Saliba‡, Rafael Bañares§, Javier Fernandez¶, Banwari Agarwal||, Rajeshwar P. Mookerjee** and Rajiv Jalan**

*IBM, Ireland. Email: msanchez, doylego, GaRanco2@ibm.ie.com

†European Foundation for the Study of Chronic Liver Failure.

Email: vicente.arroyo, marco.pavesi, elisabet.garcia@efclif.com

‡Hôpital Paul Brousse, Villejuif, France. Email: faouzi.saliba@pbr.aphp.fr

§ Hospital General Universitario Gregorio Marañón, Madrid, Spain Email: rbanares@ucm.es

¶ Hospital Clinic Barcelona, Spain. Email: jfdez@clinic.catt

|| Institute for Liver Disease Health Royal Free Hospital, London, UK. Email: banwari.agarwal@nhs.net

** University College London. Email: r.mookerjee, r.jalan@ucl.ac.uk,

Abstract—The incidence of chronic liver disease has increased in Europe and can lead to Acute on Chronic Liver Failure (ACLF) which is associated with high levels of mortality due to multisystem organ failure. The characteristics of the ACLF patients can change very rapidly within a short period of time. Continuous assessment of their recovery status is critical for clinicians to adjust and deliver effective treatment. The aim of this paper is to validate the usefulness of a data preparation approach by combining different criteria to replace missing values, balance target-class variables, select useful patient characteristics and optimise hyperparameters of machine learning models for the prediction of ACLF associated mortality rates. A key step in the data preparation is a feature selection Mutual Information (MI) based multivariate approach to build smaller, and yet equally and in some cases more informative, subsets of patient characteristics than those frequently proposed for the prediction of mortality, from patients with ACLF in the CANONIC dataset. The usefulness of the data preparation approach proposed to predict mortality was evaluated by training the XGBoost and Logistic Regression models with the prepared data. Evaluations of the models trained using a test set provided evidence of an overall high accuracy in the prediction of the mortality rates of patients for days after their diagnosis, and in some cases even higher when reduced and more informative subsets of patient characteristics were found.

I. INTRODUCTION

The incidence of chronic liver injury has increased and it can lead to ACLF, a condition associated with rapid deterioration in clinical status (usually four weeks of onset) and often leading to multi-organ failure and high mortality. There are an estimated 170,000 patients dying from ACLF each year in Europe alone with limited treatments options, including liver transplantation, which is restricted by a short time window to intervene, progression of organ failure and limitations in organ availability.

Efforts, within the medical field, are constantly carried out to minimise the mortality rates by improving the current dialysis procedures for treating patients. The achievement of such improvements is part of the objectives of the ALIVER project [1], on which this research work is based. Other

objectives include to reach a more accurate diagnosis and prognosis of ACLF to help guide better the dialysis therapies and improve patients' overall quality of health, allowing them to live better and wait longer for a liver transplantation.

The therapies need to be provided within specific time points to the patients for them to improve their response to the dialysis treatments. The optimal therapy delivery time frame for patients could be better determined by providing future overall information on organ failures and decompensations; or in any case on their possible future recovery paths captured in the prediction of their mortality rates. The accurate prediction of these patients' mortality is the core objective of this work. The prediction is carried out by training Machine Learning (ML) models using mortality historical related data of patients from the CANONIC dataset, for which detailed information on its variables can be found in [3]. The predictions aim to provide clinicians with additional information to help assess better the likelihood of ACLF mortality. The assessment is not trivial, as even when patients show recovery sudden infections or organ failures occur.

The incidence of ACLF mortality is very high within the first three months after its diagnosis. Around 30% of patients show health complications and ultimately die, even those that have been discharged from the hospital with a reasonably improved health condition can, frequently, present deterioration, as they cannot receive the same quality of care as in the hospital. The frequency of hospital readmission within three months of discharge is of the order of 40% as reported in several studies [5]. Also when a complication is severe enough and not treated when required it can lead to a higher severity of ACLF, this is denoted by the levels one, two and three, which correspond to the number of organs failing in the patients, and therefore to a considerable increase in their mortality rate.

The accurate prediction of mortality rates would help to determine patients that might require an adjustment on their dialysis therapy, or at least require a closer form of monitoring to help their likely futile course by offering

palliative options accordingly.

To boost the predicted accuracy of the mortality rates required, a data preparation approach is proposed in this paper that addresses important inconsistencies in the CANONIC data, gathered from a selected subset of ACLF patients for this work. The first inconsistency addressed is the missing patient observation values, which increase considerably in number after patients die and are therefore unable to contribute with more data records. The second inconsistency addressed is the imbalance of the classes which is rather severe especially during the initial days of prediction, right after the patients diagnosis of ACLF when too few positive data records, corresponding to dead patients, can be found. A third inconsistency addressed is the selection of data variables, called patient characteristics, that are essential in the accurate prediction of mortality. The selection is performed by eliminating non-informative variables and by keeping only subsets that are found highly informative for the accurate prediction.

The selection of the variables that correspond to patient characteristics is particularly useful in reducing the data gathering effort, and simplifies the further steps taken in the data preparation and in the building of the ML predicting models. Also the small subset variable selection helps with the overall interpretability and understandability of the patient characteristics deeply correlated in the development and progression of the ACLF condition of patients, which could help clinicians to advance steps for setting a standard criterion for its diagnosis. Previous works have been carried out to predict the ACLF different levels of intensity and the mortality rates of patients, by proposing scores, such as the SOFA and CLIF-SOFA to associate potential factors and provide insights in the development of ACLF [8]. Some of these previous works, in some cases, use larger amounts of patients characteristics than those considered in this work. Such a reduction on the data used in accurately predicting ACLF related mortality rates is the main contribution of this paper, with the data preparation approach proposed.

In the following section a recent state-of-the-art work is presented that includes relevant ideas for the improvement of care and outcome of ACLF patients. The ideas proposed are based on acquiring patients' data of different medical symptoms and signs at different stages of progression of the ACLF to mainly reduce hospital readmission rates.

II. STATE OF THE ART

Current results indicate that there is still work to be done to provide clearer explanations of ACLF developments and progressions in patients presenting different scenarios in terms of acuteness and major determinants of short term mortality. Efforts in providing such explanations have taken place in different liver disease hospitals from Europe and the United States (US).

In the US the hospitalisation costs have increased from around five billion dollars in 2001 to ten billion dollars in 2011. In[2] it is discussed how telemedicine and other novel

health remote monitoring strategies could provide better healthcare to more patients, keeping the costs low.

In [3] a diagnostic criterion is proposed based on the CANONIC dataset to determine different grades of ACLF, to predict the need for liver transplantation and the progression of severity in patients, and also the progression and response to treatments over time. This work also provides an assessment of the statistical significance of the different variables' clinical usefulness at different points in time to create a relational basis for ACLF therapy.

In [4], [5] multicentre statistical studies shows that a total of 53% of patients required hospital readmissions in a three-month time span. The work in [5] shows that a better prediction of the readmissions requires relevant variables in diagnosing ACLF, which might help recognizing tangible areas of intervention. This work used multivariate LR models and a new proposed score based on admission and discharge variables to predict hospital readmissions.

The predictions of hospital readmissions were performed using two different datasets. The first model was created using admission variables, and the second model was created using variables generated within the hospital discharged time frame. The results showed that 30% of the readmissions could not be predicted.

In [6], clinical characteristics and laboratory parameters are used within the score called CLIF-C AD to provide an increased mortality accuracy prediction, primarily within the three-month time span. The results in this work showed that this score outperforms, by 8% to 15%, other scores such as the one called Model for End-Stage Disease. The CLIF-C AD score was also used as a criterion to determine patients with a high and low risk of health deterioration, to consider patients that could be discharged early.

Another score is proposed in [7] for the prediction of ACLF patient prognosis. The score was defined simpler than other scores by being based on organ failure variables, and also including age and white blood cell count. The score proposed provided a significantly higher predictive accuracy than other proposed scores such as the MELD, MELD-Na, and CPs.

In [8] the current definition and potential causes of ACLF are discussed based on proposing variables in building more accurate predictive models. To assess the variable usefulness univariate statistical relationships were determined using the chi2, student t-test and ANOVA. The relationships aimed to determine potential factors in the development of ACLF and in the prediction of mortality rates. The predictive models were trained by implementing a multivariate stepwise forward method using statistical significance.

Other efforts are presented in [9] by considering a Remote Monitoring System for the improvement of healthcare of outpatients or patients after being discharged from the hospital. It is included the evaluation over time of the efficacy and financial sustainability of the care management check-ups on a regular basis. The evaluation consisted of assessing the differences in outcome between the patients provided with specific care management check-up and patients pro-

vided with a standard outpatients care. It was concluded the existence of advantages in monitoring the patients' levels of recovery in a continuous and remote way by reducing the hospital readmission rates. In this work efforts to help reduce the hospital readmission rates were carried out, but concluded much research is required before accurate data driven recommendations can impact significantly to help alleviate liver conditions.

In the following section the steps followed in this work are presented. Information on the CANONIC data is given, including signs and indications gathered from patients, and the steps followed to predict the mortality rates across specific time spans within the first month after ACLF diagnosis.

III. METHODOLOGY

The CANONIC data used can be organized in two categories of clinical variables the first one corresponding to clinical and biomarkers characteristics gathered under standardized laboratory techniques [8]. The second category contains only biomarker variables obtained using laboratory techniques of exploratory kind [8]. Three different groups of variables were built, using the two mentioned categories, in order to explore their predictive power in predicting mortality rates.

The first and second group corresponds to the first and second categories mentioned, and are called Non-Exploratory (NE) and Exploratory (E), respectively. The third group called Both (B) contains both the first and second group. Results on the predictive power provided by these three groups can be found in the next section in Table III.

A high number of missing values can be found in all of the three groups considered, increasing these even more in the latter days of prediction. If the data instances are simply dismissed where missing values are found, there would not be enough data to carry out the training and testing of such ML models. For this reason and as a first step of the data preparation approach proposed, two methodologies for replacing the missing values with data were explored and tested, given their relative simplicity. The first method simply replaced the missing values with the Mode (M) value of the variable where a missing value is found. The second method replaces the missing values with a number generated from a Random Normal Distribution (RND) that has the same mean and Standard Deviation (STD) of the variable where the missing value is found.

After the missing values are replaced, and as a second step the dataset is randomly split into train and test sets. The 80% of the data instances were assigned to the train set and the remaining 20% to the test set.

Using exclusively the training set a third step in the data preparation approach process, is the multivariate Feature Selection (FS) using the MI as criterion, which is a very well known similarity information measure. And similarly using only the training set a fourth and last step is to balance the number of positive and negative class data labels for each of the prediction days. After the dataset is prepared, a number of hyperparameter optimization ML approaches are explored

to determine a useful combination of parameters for training the predictive models.

A. *The CANONIC dataset*

The patient characteristics used for ML training and testing were extracted from the CANONIC dataset, from across the first 22 days after the hospital inclusion day. The hospital inclusion day being the first day the patients arrived to hospital and therefore when data started being gathered. The initial data gathered are called inclusion day. Data collected two days after the inclusion day are called "day 1" and "day 2" respectively. The data gathered between day 3 and day 7 are called the "day 3" data. The data gathered between day 8 and day 14 are called "day 8" data. The data gathered between the day 15 and day 21 are marked as "day 15". From day 22 to day 28 the data are marked as "day 22". All the data gathered between day 1 and day 22 was considered within the proposed data preparation approach for the prediction of mortality rates at day 29.

The mortality target variables, used to train the predictive models, for the prediction days is organised in the CANONIC data as follows. The mortality target variable at day 1 contains binary information of all the dead and living patients at this day. The data instances corresponding to the living patients are labelled with a zero value, and the dead ones labelled with the number one. The mortality target variable varies for each of the prediction days as the number of dead and living patients changes depending of the prediction day.

Both the targets and the patient characteristics for each of the days were used to train the ML models in a supervised way. The training and testing data used to build the predictive models for day 1 had data only from the inclusion day. For the remaining days, all the previous data instances available before a particular prediction day were considered for the training and testing of the models. For instance, for training and testing the models to predict the mortality at day 3 all the patient information available from inclusion day to day 2 was considered.

The technical approach for the selection of useful patient characteristics, using the MI, for each of the days of prediction mentioned before is described in the next section.

B. *The MI as a feature selection criterion*

The MI is a popular technique and has been used extensively across the probability and information theory fields [11]. The MI can determine linear and non-linear relationships amongst variables or groups variables and provides a quantification for the information they share.

Despite the advantages in using MI its calculation can be negatively impacted when there is very little variance in the data, which happens when many repeated or missing values are present, caused by inconsistencies in its gathering process, or reduced number of data instances [12], [13]. Another disadvantage of the MI is that its multivariate calculation becomes computationally very expensive in higher multi-dimensional spaces and therefore its usage is often restricted mainly to univariate FS analysis [14].

In the following two sections 3.3 and 3.4 a description of these two methodologies, univariate and multivariate, is provided to highlight the similarities and dissimilarities between both methods. And to understand better the potential of the multivariate analysis in describing more complex phenomena than its univariate counterpart. It is important to mention at this point that the multivariate approach, in the data used, provided higher AUC performance while providing smaller subsets of patient characteristics than the univariate method without jeopardising the level of prediction accuracy when using the whole set of characteristics. Therefore only the results obtained when selecting characteristics using the multivariate approach are reported in this work.

C. Univariate feature selection

The univariate FS analysis is less computationally expensive than the multivariate approach. This is mainly as the univariate approach does not use combinatorial search strategies to determine the useful features. A disadvantage of the univariate analysis is that it does not consider the relationships among groups of variables and how these can lead to reduced amount of data with higher predictive power. Which in this work, could lead to not capturing the ACLF heterogeneous complexity provided by the different patient characteristics.

Despite the univariate methods being less sophisticated and tending to provide less accurate predictive results than the multivariate approaches, they are sometimes preferred when providing enough performance and given their lower computational cost and scalability [15].

D. Multivariate feature selection

The MI information and the Sequential Forward Selection (SFS) approach were combined in this work for the evaluation and search of useful combinations of patient characteristics in accurately predicting mortality rates.

For the calculation of the multivariate MI in this work, the package called Gaussian-Copula Mutual Information (GCMi) version 3.0 was used under the Python GNU General Public License [16]. This package is used for the calculation of the MI between a multidimensional data input (patient characteristics) and a vector output (mortality patient label) with discrete values using a Gaussian mixture model [16]. The GCMi package calculates a multidimensional MI assuming the data variables are normally distributed.

The SFS algorithm was implemented from scratch and embedded within the GCMi for the multivariate calculation of the MI with the aim of capturing relevant relationships between the mortality CANONIC variable and the different groups of patient characteristics explored.

The multivariate FS approach allows for the measurement of the patient characteristics usefulness considering their group-relationships to the mortality variable, which aims to determine higher relevance in predictability of the mortality using smaller groups of patient characteristics.

If after considering a new patient characteristic for evaluation of the MI, following the SFS approach, a non-positive

covariance matrix is calculated within the evaluation of the MI algorithm, then this new patient characteristic is considered redundant, which means that the new patient characteristic can be represented as a linear combination of other variables that are part of the group that have been already selected. When this is the case the new patient characteristic is dismissed and not considered part of the subset of useful patient characteristics in the prediction of mortality [13], [17].

When the dimensionality of a patients' characteristics subset grows in number, it is not rare for the GCMi algorithm to calculate more of the non-positive covariance matrices. This type of matrices, especially in the latter days of prediction, result when an increased amount of information from the same patient characteristic types is collected, with some of it becoming increasingly redundant. The patient characteristics that cause the calculation of the non-positive covariance matrices are considered non-essential in the prediction of mortality, and so dismissed before training the ML models.

The patient characteristics that combined provided higher MI quantities were considered highly informative and useful in the prediction of mortality. As opposed to other combinations that provided lower levels of MI, and therefore are considered less informative and are least useful in the prediction of mortality.

In the present and previous subsections it is summarized the main differences between the multivariate and the univariate FS approaches; and highlight the potential advantages in using the former over the latter approach for the purposes of this work. In the following sections the next steps after FS in the data preparation approach are described.

E. The class data balance and the hyperparameter optimization

After the selection of the patient characteristics two class balanced approaches were explored and applied to the training data at the different prediction days to alleviate the class imbalance problem and to avoid the overfitting of the models during the training phase. The class imbalance is generated by the low number of positive labels (dead patients) and the high level of negative labels (living patients) present especially in the training data used for the prediction of the first two days. The class balance approach used for these two days was the Random Over Sampler [18]. For the remaining days of prediction from day 3 to day 29 the class imbalance is not as severe as in the first two days, the SMOTEEN class balance approach was used as it provided higher Areas Under the Curve (AUCs) results than other methodologies explored within the imbalanced-learn package [19].

It is important to remark that the test data partition was not balanced using any methodology, and that it was exclusively the subject of a pre-processing data analysis for replacing of its missing values. The training data partition, contrary to the test data partition, and once balanced, was used to fine tune the hyperparameters of the ML models compared in performance, the LR and the XGBoost(XGB).

Two approaches for hyperparameter optimization were tested in this work, Randomized [20] and Bayes Search Cross Validation [11]. The Grid search was not considered given its rather high computational complexity. No statistical comparison of the hyperparameters optimization techniques alone is provided in this work, as it is beyond its objectives

F. Building of the machine learning models

The two ML model types, the LR and the XGB, were tested in the prediction of mortality for day 1, 2, 3, 15, 22 and day 29. A total of 112 different models (see Table III) were built for each of the two ML types and for each of the configurations for data preparation explored. The models built for each prediction day were compared in performance for the different patient characteristics selected. The first criterion of selection uses the MI multivariate FS approach to determine useful variables from a subset of patient characteristics that exclusively contain NE clinical variables. The second criterion similarly uses the MI multivariate FS approach, but in this case, it selects variables from a subset that only contains E clinical group type variables. The third criterion used the MI multivariate FS approach to select useful variables from both (B) E and NE type of patient characteristics. And lastly models were built without selecting patient characteristics at all (none), this is all the variables available per day were used.

The models built with the selected patient characteristics were also compared to the models built using all the patient characteristics available per day. The results provided by the different ML models built using the different patient characteristics can be found in the following section.

IV. RESULTS AND COMPARISONS

The ML predictive models built for the present work usually predicted the negative mortality class without any trouble as there were plenty of data instances of this type (patients alive), especially in the initial days. For this reason, attention was given to the patient characteristics that could provide higher levels of recall (dead patients) and considered as the most useful in the prediction of mortality.

Table I shows the percentage of missing values increased across days. More than half of the variables contain missing values from day 1, and most of the instances have at least one missing value. Also from day 3 all the instances are incomplete. By day 29 only a small fraction of 8.82% of the variables is complete of the 255 instances available.

In Table II the number of patient characteristics selected when using the NE and E clinical type and both (B) as criterion are presented. Also contains the quantities selected, when these clinical type groups are combined with the MI FS approach for the selection of compact subset of patient characteristics, when both the Median and the RND were used to replace the missing values.

In Figure 1 information regarding the data class imbalance across the days of prediction is provided. It can be seen how the imbalance is very severe in the initial days when no many patients have passed away.

TABLE I
PERCENTAGE OF DATA AVAILABLE PER DAY

Prediction day	Data available in columns %	Data Available in records %	Number of missing values
1	40.62	0.39	6630
2	32.14	0.39	8978
3	24.19	0	11594
8	16.73	0	20383
15	12.89	0	31293
22	10.49	0	46566
29	8.82	0	62632

TABLE II
NUMBER OF PATIENT CHARACTERISTICS USED EACH DAY (M-RND)

Pred. day	Total Num	NE	E	NE+MI	E+MI	B+MI
1	97	58	40	4-5	5-4	6-4
2	145	103	40	11-10	7-9	8-7
3	193	148	40	18-17	19-16	18-19
8	278	193	77	42-44	36-41	39-42
15	361	238	112	48-50	26-48	48-50
22	444	283	147	60-54	40-55	61-36
29	527	328	182	62-60	17-58	20-60

The class imbalance is less severe in the latter days of prediction when as the ACLF condition is progressed more patients have died. The accuracy of the ML models was evaluated using the AUC as performance criterion. In Table III it can be seen that both the LOG and XGB ML models achieved different levels of performance across the different days of prediction, given that in many of cases for each day different criteria were used for the preparation of the data. The accuracy provided for each of the data preparation configurations is compared in Table III. Every combination tested results and comparisons can be found in the first column of Table III. A intra-configuration comparison was carried out as follows.

The training and testing of the ML models were repeated 10 times for statistical comparison purposes. The mean of the highest and second highest AUC performances in each configuration were compared. The statistical significance of the comparison between the best two configuration models is quantified, and the results were provided in the P-intra-Conf (p-value) .

Fig. 1. Class Imbalance across days

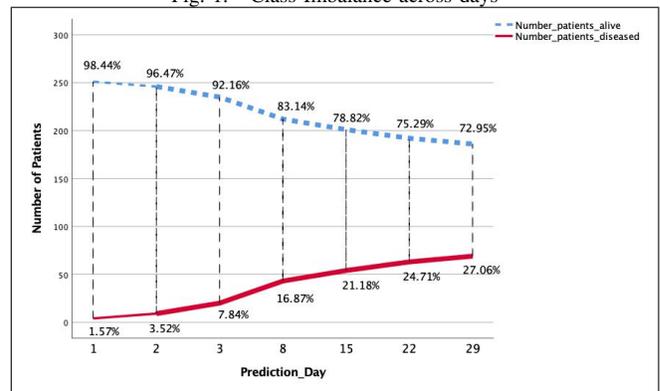
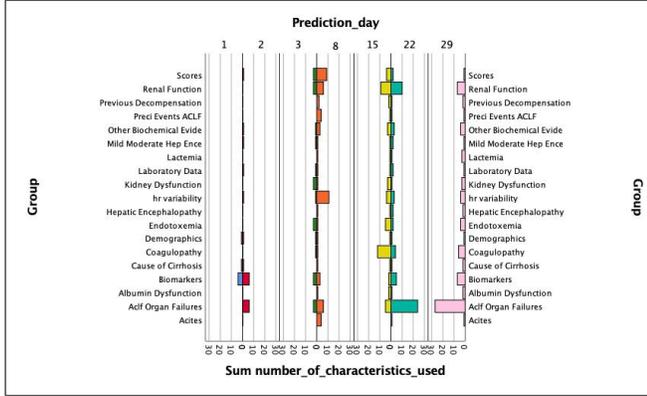


Fig. 2. Usage levels of the different groups of patient characteristics



In Table III an inter-configuration (not intra-configuration) data preparation comparison is carried out to determine the one that provided the best ML model prediction performance. The comparison is carried out as follows. The mean AUC of the two best performer models across configurations were compared per day and quantified in the P-inter-Conf row of Table III. The comparison provides information on which is the best data preparation configuration per day in the prediction of mortality. When the p-value is <0.05 the top two most accurate AUC results for a particular day are considered to be statistically significant, when $p>0.05$ these two top results are considered equally accurate and therefore the one that corresponds to the configuration that uses less patients characteristics should be used. All the statistical significance analyses were calculated using the Mann-Whitney U test.

In Figure 2 it is shown the most used groups of patient characteristics within the best data preparation configurations, these are C2.3, C1.1, C4.3, C4.3, C1.1, C4.1 and C4.1 for day 1, 2, 3, 8, 15, 22 and 29 respectively, The horizontal Sum axis quantifies within histograms the number of patient characteristics used from each of the groups of patient characteristics, ranging from 0 to above 20.

V. CONCLUSION

In the prosed approach several configurations were explored and tested to prepare the CANONIC dataset used. Each of these configurations use different techniques to provide different AUC accuracies across the days as follows.

For the replacement of missing values, the RND criterion appeared in the best configurations found for the days 3, 8, 22 and 29. While the Mode criterion appeared only for the days 1, 2 and 15. In a second step, when filtering and selecting the patient characteristics, the NE+MI criteria appeared in the best configurations found for the days 1, 15, 22 and 29. For the remaining days the B+MI criteria appeared in the best configurations.

In a third step when balancing the class variable, the Radom Over Sampler Technique was part of the best configurations for the day 1 and 2. And the SMOTEEN was part of the best configurations for the remaining days. In a fourth step to fine-tune the ML hyperparameters, the

Bayesian Search cross-validation approach was used in all of the prediction days. In a fifth step we evaluated the impact that the prepared data had in the ML models performance. The XGB model performed better for the majority of the prediction days, being the day 1 and 15 the only days in which the LOG model overcame the XGB model.

In the day 1 the configuration C2.3 provided a AUC mean very high of 0.97, despite this best high accuracy, one must consider that the test set for this first day had a very severe data class imbalance, with very few positive instances to evaluate. Therefore, future tests should be made over more cases of positive instances.

In the day 2 a best AUC mean of 0.86 was provided by the configuration C2.3. The inter-configuration p-value (>0.05) indicates that there is no statistical significance with the second-best configuration C1.1 which, only used 11 patient characteristics therefore, should be considered as the best configuration of the two. In the day 3 a best AUC mean of 0.9 was provided by C4.3. For this day 3 also there was not statistical significance with the second best configuration C1.4-none which uses 193 characteristics. Since C4.3 used only 19 patient characteristics this configuration remains as the best of the two. For the day 8 the top two best result, were generated by the configurations C4.1 and C4.3, which were not statistically significant ($p>0.05$). The C4.3 used 42 patient characteristics and can be considered as the best configuration of the two, since it used less than C4.1.

In the day 15 the best AUC mean found was 0.81 which was generated by the configuration C1.1. The top two AUC results for this day were not statistically significant either, but since C1.1NE used fewer patient characteristics (only 48), it can still be considered the best of the two. For the day 22 the best mean AUC was 0.827 generated by C4.1.

And finally for day 29 the best AUC mean obtained was 0.758 generated by C4.4-none. The intra configuration p-value (>0.05) indicated there was not statistical significance with the second best mean AUC (0.667) provided by C4.1NE which then can be considered as the best configuration of the two since the latter uses less patient characteristics for the prediction of mortality.

The day 29 had the lowest AUC prediction rate, but also this day had the highest percentage of missing values and the highest number of patient characteristics to select. And since replacing the data artificially invariably affects the quality of the data, the MI information calculation can also become less accurate, impacting ultimately in the usefulness of the patient characteristics selected and used in building the ML mortality predictive models. Overall, the multivariate MI FS approach showed to be useful in decreasing considerably the amount of patient characteristics used for ML training. Reducing the data dimensionality required can be very useful since information within the healthcare space in many cases is expensive or simply unfeasible to acquire.

The best configurations of techniques found in this work will be considered in future exploratory analyses in the prediction of ACLF related mortality. These configurations will be challenged using new exploratory model architectures

TABLE III

PERFORMANCE COMPARISONS OF THE PERFORMANCE ACCURACY (AUC) GENERATED BY THE DATA PREPARATION CONFIGURATIONS FOR THE PREDICTION OF ACLF RELATED MORTALITY USING THE LR AND XGB ML MODELS

Conf1 LR-M	day1	std1	day2	std2	day3	std3	day8	std8	day15	std15	day22	std22	day29	std29
C1.1 NE+MI	0.371	0.007	0.807	0.114	0.756	0.095	0.723	0.069	0.814	0.039	0.795	0.012	0.641	0.024
C1.2 E+MI	0.782	0.004	0.791	0.007	0.674	0.036	0.591	0.102	0.501	0.026	0.550	0.041	0.529	0.000
C1.3 B+MI	0.767	0.009	0.796	0.000	0.561	0.009	0.644	0.014	0.477	0.056	0.716	0.011	0.558	0.032
C1.4 none	0.456	0.005	0.655	0.008	0.831	0.078	0.842	0.031	0.831	0.005	0.789	0.052	0.605	0.003
P-Intra-Conf	0.001		0.057		0.395		0.351		0.351		0.351		0.351	
Conf2 XGB-M	day1	std1	day2	std2	day3	std3	day8	std8	day15	std15	day22	std22	day29	std29
C2.1 NE+MI	0.474	0.009	0.488	0.009	0.483	0.014	0.681	0.084	0.717	0.063	0.742	0.038	0.678	0.041
C2.2 E+MI	0.782	0.144	0.836	0.000	0.569	0.099	0.577	0.072	0.572	0.031	0.51190	0.04	0.476	0.023
C2.3 B+MI	0.970	0.024	0.863	0.052	0.503	0.031	0.629	0.061	0.538	0.112	0.664	0.046	0.625	0.038
C2.4 none	0.484	0.014	0.499	0.004	0.533	0.066	0.739	0.121	0.744	0.053	0.733	0.042	0.696	0.027
P-Intra-Conf	3E-05		7E-05		7E-05		7E-05		7E-05		7E-05		7E-05	
Conf3 LR-RND	day1	std1	day2	std2	day3	std3	day8	std8	day15	std15	day22	std22	day29	std29
C3.1 NE+MI	0.39	0.0	0.417	0.021	0.756	0.052	0.748	0.084	0.72	0.059	0.588	0.012	0.657	0.042
C3.2 E+MI	0.354	0.013	0.581	0.012	0.668	0.03	0.616	0.086	0.549	0.043	0.552	0.023	0.39	0.053
C3.3 B+MI	0.342	0.006	0.653	0.007	0.708	0.072	0.694	0.082	0.573	0.017	0.596	0.017	0.4	0.019
C3.4 none	0.431	0.003	0.699	0.0	0.71	0.057	0.622	0.078	0.639	0.006	0.681	0.034	0.57	0.02
P-Intra-Conf	1.E-05		2.E-05		0.029		0.284		0.284		0.284		0.284	
Conf4 XGB-RND	day1	std1	day2	std2	day3	std3	day8	std8	day15	std15	day22	std22	day29	std29
C4.1 NE+MI	0.654	0.248	0.48	0.014	0.651	0.084	0.867	0.016	0.647	0.07	0.827	0.038	0.667	0.028
C4.2 E+MI	0.486	0.027	0.577	0.112	0.608	0.056	0.643	0.036	0.565	0.03	0.51	0.045	0.5	0.037
C4.3 B+MI	0.496	0.005	0.493	0.01	0.9	0.052	0.874	0.021	0.618	0.058	0.638	0.047	0.494	0.043
C4.4 none	0.488	0.012	0.522	0.07	0.709	0.05	0.83	0.049	0.7	0.067	0.82	0.046	0.758	0.018
P-Intra-Conf	0.485		0.485		8E-05		0.485		0.485		0.485		0.485	
P-Inter-Conf	4E-05		0.136		0.191		0.02		3.8E-4		0.012		7E-04	

and new generated data from the ALIVER project.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 733057.



DISCLAIMER

Any dissemination of results reflects the authors' view only. The European Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

- [1] <http://aliver.info/>
- [2] Chirapongsathorn S, Talwalkar J, Kamath P (2016) Strategies to Reduce Hospital Readmissions. *Semin Liver Dis* 36:161–166 . doi: 10.1055/s-0036-1583196
- [3] Gustot T, Fernandez J, Garcia E, et al (2015) Clinical Course of acute-on-chronic liver failure syndrome and effects on prognosis.
- [4] Bajaj JS, O JG, Rajender Reddy K, et al (2012) Second Infections Independently Increase Mortality in Hospitalized Cirrhotic Patients: The NACSELD Experience. *Hepatology* 56:2328–2335 . doi: 10.1002/hep.25947
- [5] Bajaj JS, Reddy KR, Tandon P, et al (2016) The 3-month readmission rate remains unacceptably high in a large North American cohort of patients with cirrhosis. *Hepatology* 64:200–208 . doi: 10.1002/hep.28414
- [6] Jalan R, Pavesi M, Saliba F, et al (2015) The CLIF Consortium Acute Decompensation score (CLIF-C ADs) for prognosis of hospitalized cirrhotic patients without acute-on-chronic liver failure. *J Hepatol* 62:831–840 . doi: 10.1016/j.jhep.2014.11.012
- [7] Jalan R, Saliba F, Pavesi M, et al (2014) Development and validation of a prognostic score to predict mortality in patients with acute-on-chronic liver failure. *J Hepatol* 61:1038–1047
- [8] Moreau R, Jalan R, Gines P, et al (2013) Acute-on-chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. *Gastroenterology* 144:1426–1437.e9 . doi: 10.1053/j.gastro.2013.02.042
- [9] Morando F, Maresio G, Piano S, et al (2013) How to improve care in outpatients with cirrhosis and ascites: A new model of care coordination by consultant hepatologists. *J Hepatol* 59:257–264 . doi: 10.1016/j.jhep.2013.03.010
- [10] Clària J, Stauber RE, Coenraad MJ, et al (2016) Systemic inflammation in decompensated cirrhosis: Characterization and role in acute-on-chronic liver failure. *Hepatology* 64:1249–1264 . doi: 10.1002/hep.28740
- [11] Ban T, Ohue M, Akiyama Y (2017) Efficient hyperparameter optimization by using Bayesian optimization for drug-target interaction prediction. *IEEE Int Conf Comput Adv Bio Med Sci ICCABS 2017–Octob:1–6* . doi: 10.1109/ICCABS.2017.8114299
- [12] <https://uk.mathworks.com/matlabcentral/fileexchange/56212-gcmi-gaussian-copula-mutual-information>
- [13] Ince RAA, Giordano BL, Kayser C, et al (2017) A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Hum Brain Mapp* 38:1541–1573 . doi: 10.1002/hbm.23471
- [14] Suto J, Oniga S, Sitar PP (2016) Comparison of wrapper and filter feature selection algorithms on human activity recognition. 2016 6th Int Conf Comput Commun Control ICCCC 2016 124–129 . doi: 10.1109/ICCC.2016.7496749
- [15] Lazar C, Taminau J, Meganck S, et al (2012) A survey on filter techniques for feature selection in gene expression microarray analysis. *IEEE/ACM Trans Comput Biol Bioinform* 9:1106–19 . doi: 10.1109/TCBB.2012.33
- [16] <https://github.com/robince/gcmi>
- [17] Liu H, Taniguchi T, Tanaka Y, et al (2015) Essential Feature Extraction of Driving Behavior Using a Deep Learning Method. *Intell Veh Symp (IV)*, 2015 IEEE 1054–1060 . doi: 10.1109/IVS.2015.7225824
- [18] Mitchell R, Frank E (2017) Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput Sci* 3:e127 . doi: 10.7717/peerj-cs.127
- [19] <https://imbalanced-learn.readthedocs.io/en/stable/>
- [20] Rojas-Dominguez A, Padierna LC, Carpio Valadez JM, et al (2017) Optimal Hyper-Parameter Tuning of SVM Classifiers with Application to Medical Diagnosis. *IEEE Access* 6:7164–7176 . doi: 10.1109/ACCESS.2017.2779794