

# Adversarial Learning on Incomplete and Imbalanced Medical Data for Robust Survival Prediction of Liver Transplant Patients

Ehsan Hallaji, Roozbeh Razavi-Far, Vasile Palade, and Mehrdad Saif.

**Final Published Version deposited by Coventry University's Repository**

**Original citation & hyperlink:**

Hallaji, E., Razavi-Far, R., Palade, V. and Saif, M., 2021. Adversarial Learning on Incomplete and Imbalanced Medical Data for Robust Survival Prediction of Liver Transplant Patients. *IEEE Access*, 9, pp.73641-73650.

<https://dx.doi.org/10.1109/ACCESS.2021.3081040>

DOI [10.1109/ACCESS.2021.3081040](https://dx.doi.org/10.1109/ACCESS.2021.3081040)

ISSN 2169-3536

Publisher: IEEE

Article published under a [Creative Commons Attribution \(CC-BY\) license](https://creativecommons.org/licenses/by/4.0/).

Received April 25, 2021, accepted May 4, 2021, date of publication May 17, 2021, date of current version May 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3081040

# Adversarial Learning on Incomplete and Imbalanced Medical Data for Robust Survival Prediction of Liver Transplant Patients

**EHSAN HALLAJI**<sup>1</sup>, (Graduate Student Member, IEEE),  
**ROOZBEH RAZAVI-FAR**<sup>1,2</sup>, (Senior Member, IEEE),  
**VASILE PALADE**<sup>3</sup>, (Senior Member, IEEE),  
**AND MEHRDAD SAIF**<sup>1</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada

<sup>2</sup>School of Computer Science, University of Windsor, Windsor, ON N9B 3P4, Canada

<sup>3</sup>Centre for Data Science, Coventry University, Coventry CV1 5FB, U.K.

Corresponding author: Ehsan Hallaji (hallaji@uwindsor.ca)

**ABSTRACT** The scarcity of liver transplants necessitates prioritizing patients based on their health condition to minimize deaths on the waiting list. Recently, machine learning methods have gained popularity for automatizing liver transplant allocation systems, which enables prompt and suitable selection of recipients. Nevertheless, raw medical data often contain complexities such as missing values and class imbalance that reduce the reliability of the constructed model. This paper aims at eliminating the respective challenges to ensure the reliability of the decision-making process. To this aim, we first propose a novel deep learning method to simultaneously eliminate these challenges and predict the patients' survival chance. Secondly, a hybrid framework is designed that contains three main modules for missing data imputation, class imbalance learning, and classification, each of which employing multiple advanced techniques for the given task. Furthermore, these two approaches are compared and evaluated using a real clinical case study. The experimental results indicate the robust and superior performance of the proposed deep learning method in terms of F-measure and area under the receiver operating characteristic curve (AUC).

**INDEX TERMS** Survival prediction, class imbalance, missing data imputation, decision-making, liver transplantation, generative adversarial networks.

## I. INTRODUCTION

Liver failure may occur suddenly or chronically under various conditions, such as viral infections, intolerance to certain medications, chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma [1]–[3]. Considering that these conditions creates a major challenge for the medical team to properly determine the liver transplantation risks for each patient, a careful decision is usually made in two steps. Initially, the donor is examined to be a perfect match for the recipient based on the compatibility of their livers. Then, recipients are prioritized based on the post-operation survival chance [4].

Various scoring systems have been developed in order to increase the patients' survival chance after liver transplantation [5]. Nonetheless, the abundance of collected data and the

limited time available for decision-making further complicates this process [6]. In recent years, as the demands for liver transplantation increased, machine learning models gained popularity for automatizing the decision-making process due to their utilizable prediction performance [7], [8].

Machine learning techniques extensively accelerate the decision-making process, minimize human errors, and enable the extraction of complex patterns from the data. However, constructing an accurate machine learning model can be very challenging from different perspectives. To begin with, class imbalance takes place when the number of patients in a diagnosis class is significantly larger than the others. This difference can create a bias in the model and decrease the accuracy of the prediction model [9], [10]. This bias is typically towards the majority class, which increases misclassification rate for the minority class. Another issue of concern is the presence of missing values [11] within the clinical

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott<sup>1</sup>.

data due to the existing time constraints. For example, such missing answers are inevitable for individuals who complete the questionnaire in the hospital.

Despite the recent advancements in machine learning, their trustworthiness may remain a concern, as the aforementioned complexities within the data can affect the reliability of the model. Here, we aim to address this problem by eliminating missing values and class imbalance in the data so that the patients' survival probability can be confidently estimated. The result if this confident estimation is an extensive improvement in the reliability of the scoring systems. To this aim, we propose a multi-task novel deep learning framework as well as a hybrid framework to handle missing values and class imbalance and predict survival chance of the recipients. The former devises adversarial learning and accomplishes all three tasks within a unified structure. Similar to most adversarial problems [12]–[14], the proposed algorithm devises two competing neural networks that create a min-max optimization problem. The latter, on the other hand, consists of multiple advanced class imbalance learning (CIL), missing data imputation, and classification techniques. These two approaches are compared and evaluated on a real-world case study in terms of F-measure and the area under the receiver operating characteristic curve (AUC).

The contribution of this paper is threefold: i) a novel multi-task deep learning algorithm is proposed for rectifying the complexities within the raw medical data and predicting the survival chance of liver transplant patients before the operation. ii) a hybrid framework is designed using the combination of multiple missing data imputation, class imbalance learning, and classification methods in different modules to tackle the aforementioned challenges. iii) a comparative study is enabled on real clinical data based on the performance of the survival chance estimation in the presence of missing values and class imbalance in the data.

The remainder of this paper is organized as follows: Section II gives the required background on liver transplantation and the challenges in using machine learning for the task at hand. Section III introduces the selected case study. Section IV explains the proposed approaches for early prediction of survival chance for liver allocation in the presence of missing data and class imbalance. The experimental results are analyzed in Section V. Finally, the paper is concluded in Section VI.

## II. BACKGROUND

Here, important scoring systems for liver allocation are initially reviewed to clarify the need for using computational models. Then, possible solutions for addressing two of the most challenging problems in automatizing this process, namely class imbalance and missing data are overviewed.

### A. SCORING SYSTEMS

Traditionally, patients were listed and prioritized based on their blood type, BMI, and medical condition (degree of disease). They were then prioritized based on three

simple tests: creatinine, bilirubin, and International Normalized Ratio (INR), which later led to the emergence of other scoring systems called MELD. Patients who received the highest score were given the top priority for transplantation. With the exacerbation of the disease, their priority score for liver transplantation increased. Using the MELD scoring system, no one could predict how long a patient would have to wait in the waiting list to receive the graft [5]. Due to the increased mortality rate in the liver transplantation waiting list, the MELD system was no longer used, and researchers were searching for more efficient ways to allocate organs to the recipients [15].

Multiple scoring systems were proposed afterwards to increase the survival chance of patients, including Donor Risk Index [16], Survival Outcome Following Liver Transplant (SOFT) [17], Donor-MELD (D-MELD) [18], and Balance of Risk [19]. However, each of these either had limitations or did not gained popularity. As a result, MELD is still used as the basic index of all decision-makings in recent liver allocation techniques in most regions of the world. Thus, the liver transplantation teams needed to remedy the shortcomings of this technique with the aid of computational models.

### B. CLASS IMBALANCE LEARNING

One of the common challenges in training computational models is evading the bias caused by class imbalance [10], [20], [21]. To address this problem, sampling is one of the techniques widely used to balance the distribution of classes [20], [22]. A popular advanced sampling technique is the Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic samples and performs oversampling in the minority class [23]. Each newly created sample is placed along the line between a chosen sample of the minority class and its nearest neighbor [24]. Reference [25] proposes another oversampling method using the real-value negative selection process. Another over-sampling method is designed in [26] by means of density peaks clustering along with heuristic filtering. Class imbalance can also be eliminated using hybrid methods that make use of sampling and algorithm level approaches [27]. Cost-sensitive learning is another category of CIL techniques that measures the costs of misclassification. These techniques typically need to specify the cost matrix based on misclassified records in the learning stage [23]. In recent years, hybrid techniques have been widely used for CIL that combine algorithm-level and data-level approaches [28]. In this work, SMOTE [29] and Adaptive Synthetic Sampling (ADASYN) [30] are used as oversampling, and NearMiss (NM) [31] is employed as an under-sampling method.

### C. MISSING DATA IMPUTATION

Missing data are very common in medical informatics, as the recorded data is usually associated with intricacies, such as inaccessibility of accurate information at the time of registration in questionnaires or lack of co-operation of the patient in the post-treatment phase. In standard data

**TABLE 1. Characteristics analyzed for recipients in the liver transplantation waiting list.**

Feature	Description	Unit/Type
Age	Age of the recipient	Numeric
Sex	Sex of the recipient	Binary
Alcoholic	If the patients is alcoholic	Binary
Cholestatic	Diagnosed with cholestatic liver disease	Binary
Hepatitis	Diagnosed with viral hepatitis	Binary
Disease	Diagnosed with other/unknown diseases	Binary
Blood	Blood type (e.g., A, B, AB, O)	Numeric
Bilirubin	Preoperative total bilirubin	(mg/dL)
Creatinine	Preoperative creatinine	(mg/dL)
INR	Preoperative INR median	Ratio
MELD	MELD median	Numeric
Event	Target class (e.g., death or transplant)	Binary

analysis, patients with incomplete information are removed from datasets [32]. However, this results in a loss of efficiency. Therefore, it is very common to estimate the incomplete parts of data, which is known as missing data imputation [33], [34]. Various techniques have been used for missing data imputation [11], [35], [36]. The easiest technique is to replace the missing value with the mean or mode of the observed data within the respective feature [37]. This can be also done w.r.t. each class of data to improve the imputation quality. Missing data imputation can be performed using more advanced approaches such as  $k$  nearest neighbors imputation (KNNI) [38], Markov models [33], least squares [39], Multiple Imputation by Chained Equations (MICE) [40], Denoising Auto Encoder (DAE) [41], MissForest (MF) [42], and Expectation Maximization Imputation (EMI) [43].

### III. CASE STUDY

The utilized dataset contains the information of a waiting list patients. These patients were registered for a liver transplantation at Mayo Clinic, Rochester, MN, US [44]. The data was collected over a 10-year period, starting from February 1990. 815 adult patients were sampled out to form the dataset. Among these patients, 37 of them were withdrawn from the list, and 76 of them were censored. The rest of the samples, which are used in our experiments, correspond to recipients whose surgery led to death or successful transplantation. 55 percent of patients are male and the rest are female.

Patients in the waiting list were analyzed w.r.t. the features listed in Table 1 before the liver transplantation. The MELD score that is used in this case study, and showed in Table 1, is formulated as follows:

$$M = 9.75 \log_e S_C + 3.78 \log_e S_B + 11.2 \log_e \text{INR} + 6.4, \quad (1)$$

where  $M$  indicates the MELD score,  $S_C$  is the serum creatinine, and  $S_B$  is the serum total bilirubin.

The dataset is mostly formed using complete patient records and contains only 0.4 percent of missing data. To simulate more challenging scenarios, we generate additional missing values using the Missing At Random (MAR) mechanism. This is mainly because, logically, some features are

more likely to be missed than the others. For instance, not every patient may remember their blood type when they complete the questionnaire. Conversely, features such as age and sex are hardly being missed. This is in contrast to other missing mechanisms, namely Missing Completely At Random (MCAR) and Missing Not At Random (MNAR). MAR missing values are induced w.r.t. the following formal definition:

Consider a complete dataset  $X^{n \times m} = \{x_1, x_2, \dots, x_n\}$ , in which a data vector  $x_i$  or a row of  $X$  is defined as  $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ . Then, some values  $x_{ij} \in X$  will be removed as  $x_{ij} = \emptyset$  to create the incomplete data  $X'$ . By employing an indicator matrix  $I = \{I_{ij} \mid I_{ij} \in \{0, 1\}\}$ ,  $1 \leq i \leq n$  and  $1 \leq j \leq m$ , complete and missing values in  $X'$  can be marked with 1 and 0, respectively. Accordingly,  $X'$  can be divided into two sub-matrices as  $X' = X_{\text{obs}} \cup X_{\text{mis}}$ , where

$$\begin{cases} X_{\text{obs}} = \left\{ x_i \mid x_i \in X' \wedge \sum_{j=1}^m I_{ij} \times x_{ij} = 1 \right\} \\ X_{\text{mis}} = \left\{ x_i \mid x_i \in X' \wedge \sum_{j=1}^m I_{ij} \times x_{ij} = 0 \right\} \end{cases} \quad (2)$$

The distribution of the induced missing values in  $X'$  can then be modeled w.r.t.  $I$ ,  $X_{\text{obs}}$ , and  $X_{\text{mis}}$ . To create MAR incompleteness in  $X_{\text{mis}}$ , a set of matrix values are selected for removal that satisfy the following condition:

$$P(I = 0 \mid X_{\text{obs}}, X_{\text{mis}}, \theta) = P(I = 0 \mid X_{\text{obs}}, \theta), \quad (3)$$

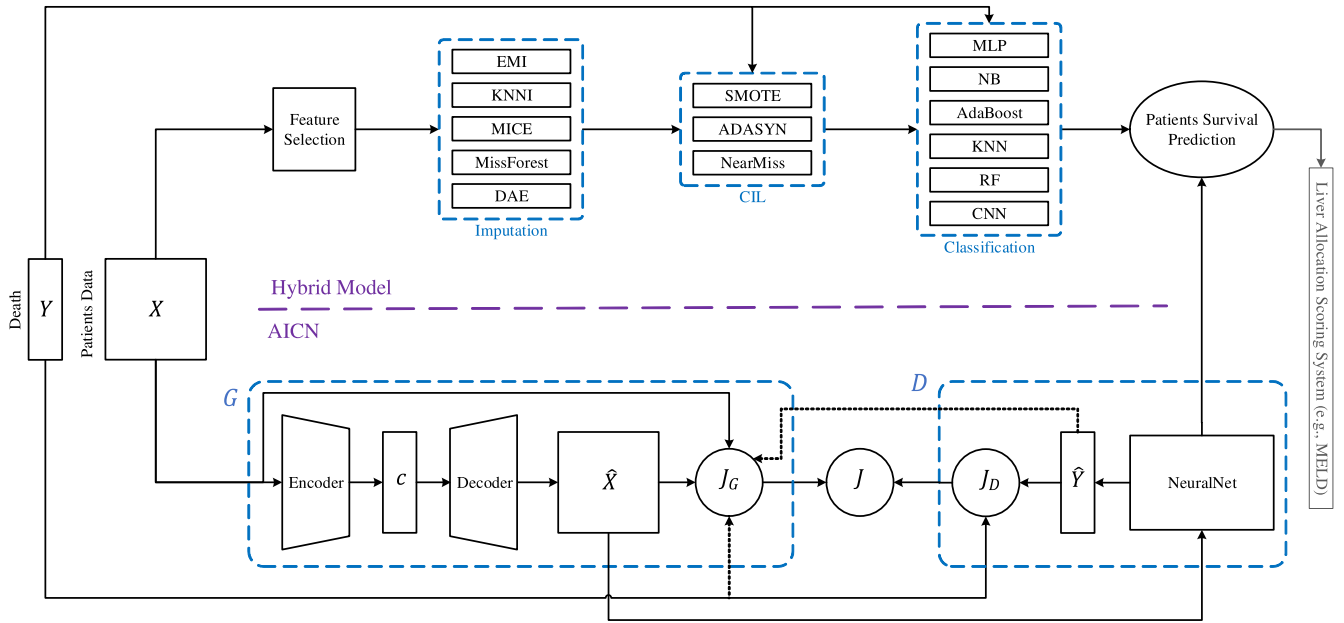
where the probability of being missed only depends on a set of parameters  $\theta$ , and the incompleteness of data may depend on the observed data as well. In this work, a normal distribution is employed to induce the missing values to which  $\theta$  corresponds.

Given that 636 of the operations were performed successful (negative class) and 66 of them resulted in death of the recipients (positive class), an imbalance ratio of 9.63 can be seen in the utilized dataset.

### IV. METHODOLOGY

Considering samples with missing values in the training set, the evaluation of the imputation cannot be fully accomplished as there are no targets available. This may result in the quality degradation of the training model. On the other hand, disregarding the incomplete samples during the training is not preferable as it does not fully use the available data.

Inspired by the famous Generative Adversarial Networks (GAN) [45], we propose a deep learning model, called Adversarial Imputation-Classification Network (AICN) (see Figure 1) to predict survival chance of patients when missing values and class imbalance exist in the training data. The proposed network is comprised of two parallel neural networks, namely generator  $G$  and discriminator  $D$  (see Figure 1) whose loss functions are tied together. The former utilizes a DAE [46] for missing value imputation while the latter performs the classification task on both complete and imputed samples. By this meaning, complete samples are directly classified by  $D$  and incomplete samples are first completed by means of  $G$  and then classified by  $D$ . Furthermore, DAE



**FIGURE 1.** Architecture of the proposed models for predicting the survival probability of patients. The upper half of the diagram shows the designed multi-step hybrid model that consists of multiple modules. The lower half of the diagram illustrates the design of the AICN algorithm, which accomplishes the same goal within an integrated scheme.

structure makes the model robust against noise in the data. To eliminate the class imbalance, a weighted neural network is utilized, in which classes are assigned with different weight w.r.t. the degree of imbalance within the training data. The algorithm is formally explained in the followings.

Given a data matrix  $X$  with  $n$  rows and  $m$  columns, we define an indicator matrix  $I^{n \times m} \in \{0, 1\}$ , in which zeros and ones indicate missing and available values of  $X$ , respectively.

**A. GENERATOR**

Initially, the missing values in each column  $j$  are replaced with  $\mu_j$ , the mean of that column. Then, the training set will go through stochastic noise corruption, where 50 percent of  $X$  is randomly replaced with zeros to create the noisy input  $\tilde{X}$  for DAE.

**1) ENCODER**

Samples in  $\tilde{X}$  are fed to the encoder of DAE as follows.  $\forall \tilde{x}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{im}) \mid \tilde{x}_i \in \tilde{X}$ :

$$\tilde{x}, \bar{z}_1, \bar{z}_2, \dots, \bar{z}_{L^E}, \quad c = \text{encoder}(\tilde{x}), \tag{4}$$

$$\bar{z}_l^* = \omega_l \cdot r_{l-1} + \beta_l, \tag{5}$$

$$\bar{z}_l = \frac{\bar{z}_l^* - \mu_l}{\sigma_l} + N_G(0, \sigma^2), \tag{6}$$

$$r_l = \max(0, \bar{z}_l), \tag{7}$$

where  $\bar{z}_l^*$  and  $\bar{z}_l$  are latent variables created at  $l$ -th layer of the encoder before and after batch normalization, respectively;  $c$  is code, or the output of encoder;  $L^E$  is the number of hidden layers in encoder;  $\omega$  and  $\beta$  are the weight and the bias vectors of the encoder;  $r$  is the output of the ReLU activation

function;  $\mu_l$  and  $\sigma_l$  are mean and standard deviation of  $\bar{z}_l$ , respectively;  $N_G$  is the Gaussian noise used for encoding.

**2) DECODER**

The output of the encoder, so called code, goes through the decoder thereafter (see Figure 1). The decoder structure is detailed in the following:

$$c, \hat{z}_1, \hat{z}_2, \dots, \hat{z}_{L^D}, \quad \hat{x} = \text{decoder}(c), \tag{8}$$

$$\hat{z}_l^* = \varpi_l \cdot r_l - 1 + \nu_l, \tag{9}$$

$$\hat{z}_l = \frac{\hat{z}_l^* - \mu_l}{\sigma_l}, \tag{10}$$

$$r_l = \max(0, \hat{z}_l), \tag{11}$$

where  $\hat{z}_l^*$  and  $\hat{z}_l$  are latent variables created at  $l$ -th layer of the decoder before and after batch normalization, respectively;  $\varpi$  and  $\nu$  are the weight and the bias vectors of the decoder;  $L^D$  is the number hidden layers in the decoder.

Given that  $Y = \{y_1, y_2, \dots, y_n\}$  is the set of class labels for  $x_i \in X$ , the cost function of the generator  $J_G$  is defined as follows:

$$J_G = -\frac{1}{n} \sum_{i=1}^n (I_i \odot X_i - I_i \odot \tilde{X}_i)^2 + \sum_{i=1}^n (I_i \wedge \hat{I}_i) \log P(\hat{y}_i = y_i \mid x_i), \tag{12}$$

where  $\odot$  and  $\wedge$  are element-wise multiplication and logical AND, respectively. Also,  $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$  is the set of predicted class labels. The first term in  $J_G$  denotes to reconstruction cost on available parts of  $X$  while the second term is the cross-entropy of the classification for incomplete

samples, which is performed by  $D$ . In other words, the second term ties the  $G$  loss function to  $D$ , which is the main element of GAN approaches.

### B. DISCRIMINATOR

Once the missing values are imputed by  $G$  and  $\hat{X}$  is attained, samples can be fed to  $D$  to perform classification. We model  $D$  as a fully connected neural network with Sigmoid activation function,  $S = \frac{1}{1+e^{-x}}$ , in the last layer for the sake of simplicity. However, more advanced structures are also compatible with this scheme as well. The discriminator aims to maximize a cross-entropy cost  $J_D$  such that:

$$J_D = \sum_{i=1}^n \log P(\hat{y}_i = y_i | x_i). \quad (13)$$

Thus, similar to most GAN approaches, the objective function  $J$  is defined as a min-max problem:

$$J = \min J_G + \max J_D. \quad (14)$$

### C. ADAPTING TO CLASS IMBALANCE

Assuming that  $\Omega = \{e_1, e_2, \dots, e_\kappa\}$  is the set of all possible events in the waiting list (e.g., death and transplant), there can be  $\kappa$  classes considered in the data. The training rate for each class is computed, as follows:

$$\forall e_i \in \Omega, 1 \leq i \leq \kappa : R_i = \frac{0.5}{\frac{\text{card}(e_i)}{n}} = \frac{n}{2 \text{card}(e_i)}, \quad (15)$$

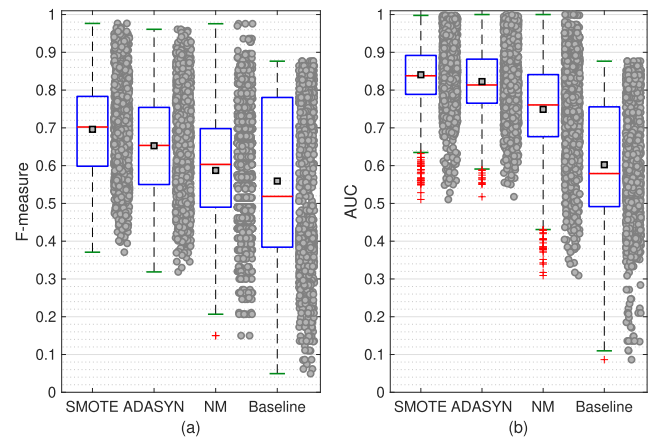
where  $\text{card}(\cdot)$  returns the cardinality. By this mean, the impact of each class on updating the weight of the neural network will be adjusted based on the population of  $e_i$  in the training data. In a balanced data set,  $R$  will be calculated as a vector of ones, which indicates equality of training impact for all classes. The estimate  $R_i$  for each class is then multiplied to the calculated loss on each class during the training, which in turn intensifies the learning on the minority class and slows the learning process on the majority class to offset the imbalance effect in the discriminator.

### D. HYBRID MODEL

A multi-step hybrid model is also designed, as illustrated in Figure 1, for the sake of comparison with AICN. This model contains four main modules, namely feature selection, imputation, CIL, and classification. The first module uses the suggested features in the case study. The imputation module employs EMI, KNNI, MICE, MissForest, and DAE methods for estimating missing values. Once the data is completed, the CIL module uses SMOTE, ADASYN, and NearMiss for eliminating the class imbalance. Finally, predictive models are constructed by means of AdaBoost [47], KNN, Random Forest (RF) [48], Naive Bias (NB), Multilayer Perceptron (MLP), and 1D Convolutional Neural Network (CNN) [49].

## V. EXPERIMENTAL RESULTS

We first explain the setting used throughout the experiments. The obtained results are compared and analyzed.



**FIGURE 2.** Obtained results for all classifiers over 10-fold cross-validation procedure grouped with the CIL techniques. AICN results are not included in this figure as it does not require CIL. Square, circles and the plus sign indicate mean, attained measurements, and outliers, respectively. The figure contains the results of all missing ratios.

### A. EXPERIMENTAL SETTING

To ensure the statistical reliability of the attained results, we devise a 10-fold stratified cross-validation scheme for all the experiments. This scheme consists of an inner and an outer loop. The outer loop keeps one of the folds for testing and the rest for the training, while the inner loop performs another cross-validation procedure on the training folds to find the optimal parameters. The inner loop employs the grid search algorithm for the parameter tuning. Furthermore, algorithms of each module are optimized by fixing those of the other modules, where the classification module is tuned after the CIL module. Moreover, the Adam optimizer [50] is used as the solver of AICN, MLP, and CNN that are all trained through 1000 epochs. Furthermore, the MLP structure is similar to the discriminator of AICN, which performs the classification task. Table 2 provides more details regarding the parameter setting of all the utilized algorithms.

Five ratios of missingness in intervals of five percent are considered with a maximum missingness of 25 percent. These missing values are generated using the MAR mechanism.

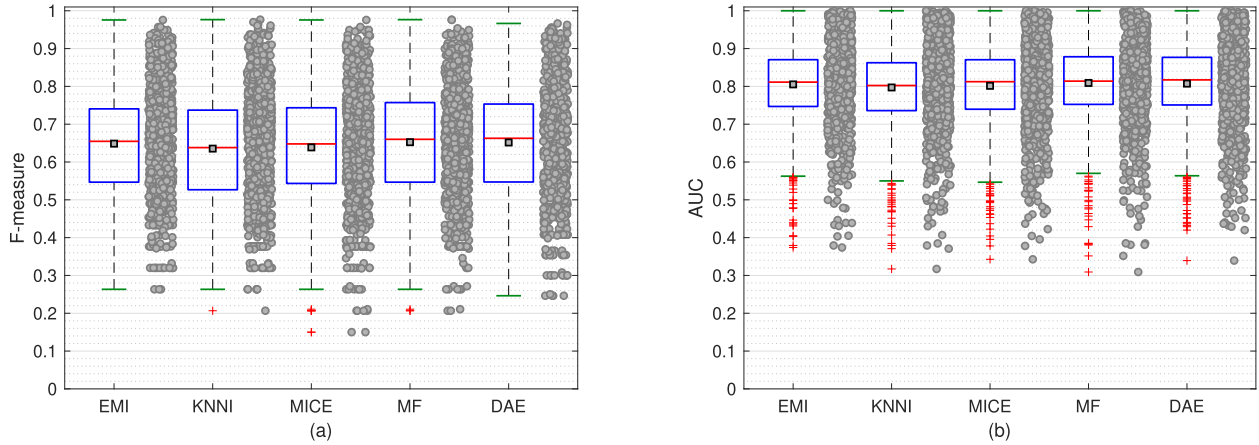
### B. ANALYSIS

A comparative analysis is performed here on the imputation, CIL, and classification techniques. We mainly conduct our analysis based on F-measure and AUC metrics.

#### 1) CLASS IMBALANCE LEARNING

To evaluate the performance of CIL methods, we assess to what extent the effect of class imbalance is eliminated by the methods w.r.t. the final AUC and F-measure.

Looking through Figure 2, it is noticeable that SMOTE results in the highest overall F-measure and AUC while ADASYN and NM are ranked as second and third in this respect (see Figure 2(b)). A higher F-measure and AUC can indicate the higher balance between the classification performance of both majority and minority classes. Furthermore,

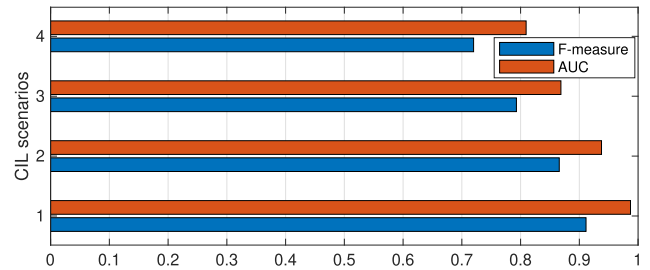


**FIGURE 3.** Classification results over 10-fold cross-validation grouped with the imputation techniques. AICN results are not included in this figure since it imputes missing data by itself. Square, circles and the plus sign indicate mean, attained measurements, and outliers, respectively. The results of all missing ratios are included in this figure.

**TABLE 2.** Parameter setting and tuning of the utilized techniques.

Algorithms	Parameter	Range/value
EMI	Max. #. iterations	2000
KNNI	#. neighbours	{1, 2, 3, 5, 7, 9}
MICE	Max. #. iterations	300
	#. imputed sets	{5, 7, 9, 11}
MissForest	Max. depth	{70, 80, 90, 100}
	#. estimators	{10, 25, 50, 100, 200}
	Max. #. samples per leaf	{3, 5, 7, 9}
	#. hidden layers	2
DAE	Learning rate	{0.001, 0.005, 0.01, 0.02}
	Batch size	{50, 100}
	Layer sizes	{100, 50} and {50, 100}
SMOTE	#. neighbours	{1, 2, 3, 5, 7, 9}
	Sampling strategy	Minority & not majority
ADASYN	#. neighbours	{1, 2, 3, 5, 7, 9}
	Sampling strategy	Minority & not majority
NearMiss	#. neighbours	{1, 2, 3, 5, 7, 9}
	Version	{1, 2, 3}
	Sampling strategy	Minority & not majority
KNN	#. neighbours	{1, 2, 3, 5, 7, 9}
RF	#. estimators	{10, 50, 100, 200}
	Max. depth	{70, 80, 90, 100}
	Min. sample per leaf	{1, 2, 3, 5, 7}
	#. hidden layers	3
MLP	Learning rate	{0.001, 0.005, 0.01, 0.02}
	Batch size	{50, 100}
	Layer sizes	{100, 100}
CNN	ID filter sizes	{3, 3}
	Dense layer size	100
	Batch size	{50, 100}
	Learning rate	{0.001, 0.005, 0.01, 0.02}
AdaBoost	Base estimator	Decision tree
	#. estimators	{50, 100, 150}
	Learning rate	{0.5, 1, 1.5}
NB	Smoothing variance	$1e-9$
	#. hidden layers	2
AICN	Batch size	{50, 100}
	Learning rate	{0.001, 0.005, 0.01, 0.02}

the baseline F-measure and AUC measurements show that the absence of CIL techniques decreases the performance, which is caused by the bias towards the majority class. Figure 2 also shows that the CIL effect on AUC is more significant.



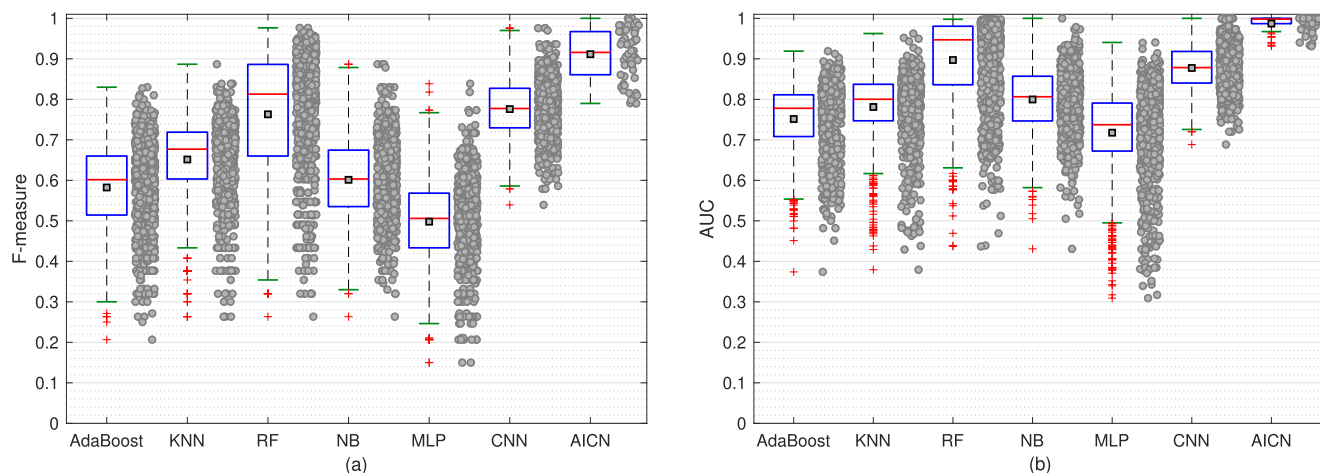
**FIGURE 4.** Performance evaluation of the imbalance handling approach used in AICN under four different scenarios. Scenario one indicates the proposed weighting approach.

To study the CIL effect on AICN, we consider four different scenarios. Firstly, we consider the proposed approach that is formulated in Equation (15). Secondly, a common approach is considered that tackles the imbalanced data by adjusting the training weight of each class as  $\text{card}(e_i)/n$ . The third scenario resembles the case when the majority class has the weight of one, the minority and majority classes are weighted with the corresponding imbalance ratio, where the majority class has a weight of one. Finally, the last scenario does not apply CIL to show the effect of imbalanced data in the absence of the mentioned approaches. Figure 4 shows the proposed method (i.e., scenario 1) outperforms others. Furthermore, it can be seen that all CIL approaches result in better F-measure and AUC compared to the last scenario, where CIL is not applied.

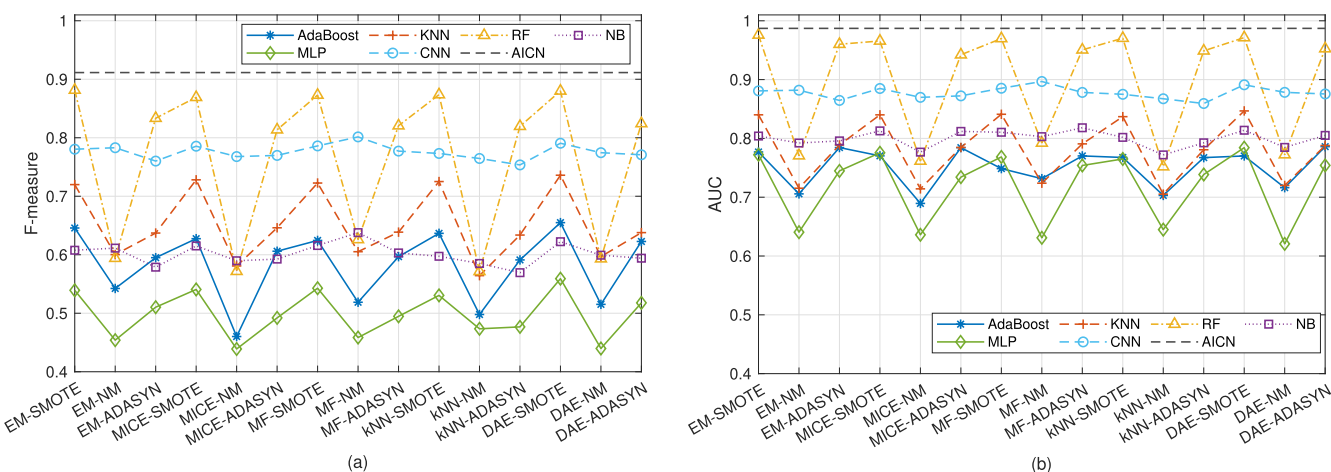
## 2) MISSING DATA IMPUTATION

In the considered scenarios, the medical data contains many multiple missing values, making it impossible to perform an evaluation based on the comparison of the imputed values with the original values. For this reason, post imputation accuracy is chosen to evaluate the data imputation techniques.

Figure 3 shows the post imputation accuracies of each imputation method for all missing ratios. Figure 3 shows



**FIGURE 5.** Classification results over 10-fold cross-validation. AICN has fewer measurements as is not combined with any imputation or CIL techniques. Squares, circles, and the plus signs indicate mean, attained measurements, and outliers, respectively. The results of all missing ratios are included in this figure.



**FIGURE 6.** Averaged accuracy and F-measure attained by AICN and different combinations of CIL, imputation, and classification algorithms. The results are averaged over all missing ratios.

that the performance of the imputation techniques are very similar. Nevertheless, MF and DAE are ranked as first and second ranks with a negligible difference. EMI, MICE and KNNI are ranked from third to the last. This comparison applies to both performance measures, AUC and F-measure, which are shown in Figure 3(a,b), respectively.

### 3) CLASSIFICATION

Here we compare the classification result of the proposed AICN, which is not combined with any other method, to other classifiers that are combined with CIL and imputation techniques. Each classifier is combined with a set of imputation and CIL techniques through different experiments, which is why the combined classifiers have more recorded measurements than AICN in Figure 5.

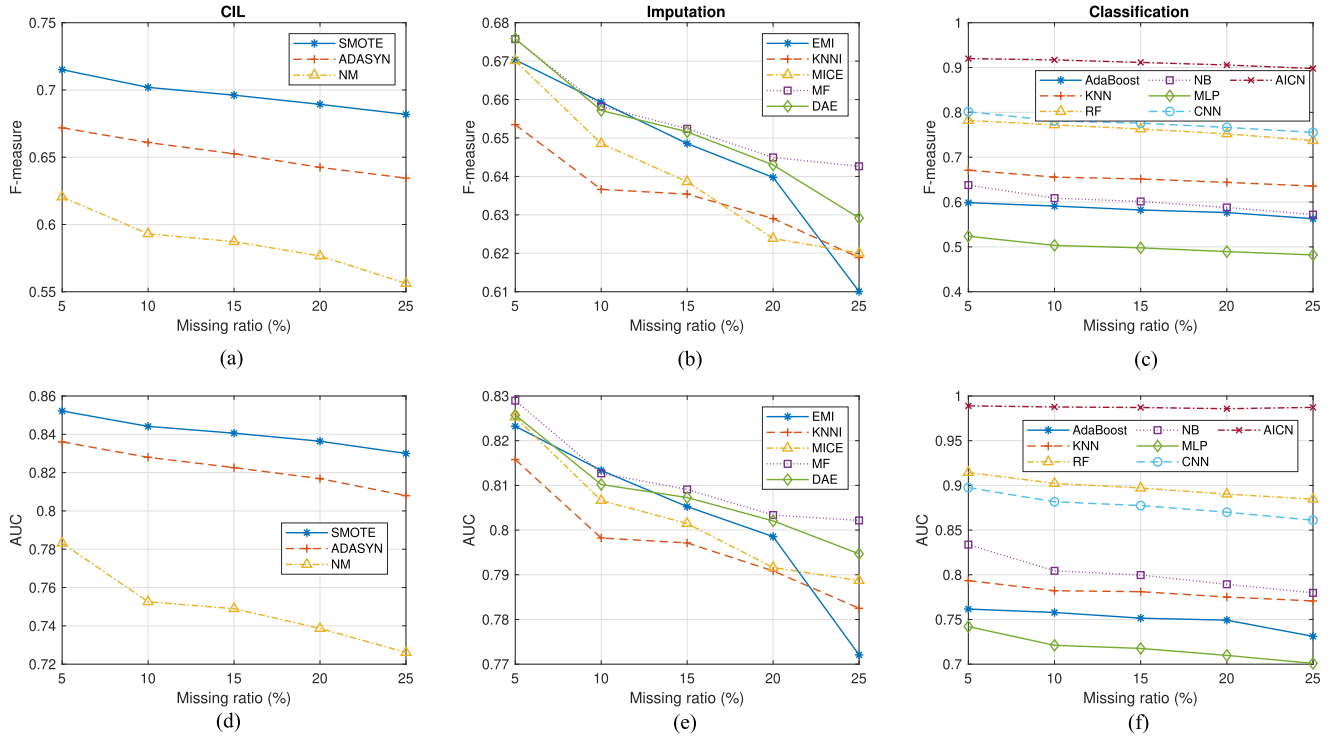
Figure 5 shows that AICN outperforms all other classifiers in terms of accuracy and F-measure. This, firstly implies that

the intrinsic imputation phase of AICN outperforms other imputation methods. Secondly, the algorithm structure is able to decrease the class imbalance effect to a large extent while creating an accurate classification model.

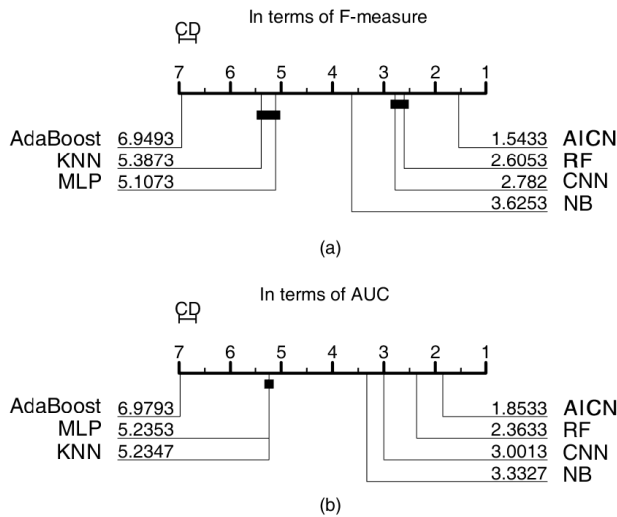
There is a noticeable difference between the attained results based on AUC and F-measure. The class imbalance has caused more bias for AdaBoost, MLP, KNN, and NB. This is while, AICN, RF, and CNN are ranked from first to third. Despite the negligible difference of RF and CNN, it seems that the latter is more stable throughout the experiments.

A more detailed comparison can be done based on Figure 6 that shows AUC and F-measure for each combination in these experiments. While the results are different for F-measure and AUC, it can be inferred that the overall ranking of the methods is similar. Here, AICN again outperforms the other methods and is ranked as first. The stability of CNN compared to RF over different combinations can be seen in this figure.





**FIGURE 7.** Average performance measures attained through each CIL, missing data imputation, and classification methods over different ratios of missing data. The results are shown in terms of accuracy and F-measure.



**FIGURE 8.** Critical difference diagram of the post-hoc Friedman test on the results of classification algorithms. All the combinations and missing ratios are considered in this test. The test is performed based on F-measure (a) and AUC (b).

However, for certain combinations, RF outperforms CNN and can be ranked as second.

In short, the best CIL performance in the hybrid model is resulted by SMOTE. As for the imputation techniques, MissForest has yielded more reliable imputed values for the CIL. Finally, AICN results in the highest accuracy and F-measure for all missing ratios.

### C. EFFECT OF MISSING RATIO

Figure 7 shows the effect of missing ratio on the performance of the CIL, missing data imputation, and classification. Generally, it can be seen that the AUC and F-measure of three tasks deteriorates almost linearly, as shown in see Figure 7. However, it seems the performance decreased more severely, from 5% to 10%. Moreover, AICN has the lowest sensitivity to the ratio of missing data in terms of both AUC and F-measure.

To study the significance of AICN, we perform a post-hoc Friedman test over the results of all ratios and combinations and compare the proposed methods with the rest of the methods. Figure 8 shows the critical difference (CD) diagram obtained from the conducted statistical test. This figure indicates that AICN significantly outperforms other methods in the hybrid model.

### VI. CONCLUSION

This paper addressed data complexities such as class imbalance and missing values within the raw medical data, while predicting the survival chance of the liver transplantation patients. This facilitates the process of training a computational model used in automated liver allocation systems and enhance their reliability. To this aim, a novel deep learner is proposed, by resorting to adversarial learning, to eliminate these challenges simultaneously within an integrated scheme. In addition, a multi-step hybrid framework is designed by employing multiple advanced techniques for CIL,

imputation, and classification within different modules. A comparison on the combinations of these techniques is performed to determine the best hybrid combination for the task at hand. Furthermore, the proposed deep learner is compared with these techniques in terms of post-imputation AUC and F-measure. Either of these models can be used to confidently predict the survival chance of patients, which in turn increases the efficiency of the scoring systems. For the sake of evaluation, a real clinical case study is selected on patients registered for liver transplantation. Finally, the experimental analysis indicates the superior performance of the proposed AICN method. Future works can extend AICN for recommender systems of different applications. Furthermore, AICN can be improved in terms of computational efficiency, and, thus, be used for big data analytics. Besides, despite the success of AICN compared to its rivals, robustness against missing value can still be improved to reach an F-measure close to 100 percent even for high missing ratios.

## REFERENCES

- [1] A. Singh, P. Nath, V. Singhal, D. Anand, Kavita, S. Verma, and T.-P. Hong, "A new clinical spectrum for the assessment of nonalcoholic fatty liver disease using intelligent methods," *IEEE Access*, vol. 8, pp. 138470–138480, 2020.
- [2] N. Li, J. Zhang, S. Wang, Y. Jiang, J. Ma, J. Ma, L. Dong, and G. Gong, "Machine learning assessment for severity of liver fibrosis for chronic HBV based on physical layer with serum markers," *IEEE Access*, vol. 7, pp. 124351–124365, 2019.
- [3] F. Hao, J. Shi, Y. Chen, X. Gu, and D. Ma, "Integrated analysis of a noncoding RNA-associated competing endogenous RNA network in non-alcoholic fatty liver disease," *IEEE Access*, vol. 8, pp. 88977–88989, 2020.
- [4] X. Muller, A. Schlegel, P. Kron, D. Eshmunov, M. Würdinger, D. Meierhofer, P.-A. Clavien, and P. Dutkowski, "Novel real-time prediction of liver graft function during hypothermic oxygenated machine perfusion before liver transplantation," *Ann. Surgery*, vol. 270, no. 5, pp. 783–790, Nov. 2019.
- [5] R. Wiesner, "MELD and PELD: Application of survival models to liver allocation," *Liver Transplantation*, vol. 7, no. 7, pp. 567–580, Jul. 2001.
- [6] M. Vidyasagar, "Machine learning methods in computational cancer biology," *Annu. Rev. Control*, vol. 43, pp. 107–127, Jan. 2017.
- [7] D. E. Schaubel, M. K. Guidinger, S. W. Biggins, J. D. Kalbfleisch, E. A. Pomfret, P. Sharma, and R. M. Merion, "Survival benefit-based deceased-donor liver allocation," *Amer. J. Transplantation*, vol. 9, no. 4p2, pp. 970–981, Apr. 2009.
- [8] M. Pérez-Ortiz, P. A. Gutiérrez, M. D. Ayllón-Terán, N. Heaton, R. Ciria, J. Briceño, and C. Hervás-Martínez, "Synthetic semi-supervised learning in imbalanced domains: Constructing a model for donor-recipient matching in liver transplantation," *Knowl.-Based Syst.*, vol. 123, pp. 75–87, May 2017.
- [9] X. Zhang and B.-G. Hu, "A new strategy of cost-free learning in the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 2872–2885, Dec. 2014.
- [10] R. Razavi-Far, M. Farajzadeh-Zanjani, B. Wang, M. Saif, and S. Chakrabarti, "Imputation-based ensemble techniques for class imbalance learning," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 1988–2001, May 2021.
- [11] E. Hallaji, R. Razavi-Far, and M. Saif, "DLIN: Deep ladder imputation network," *IEEE Trans. Cybern.*, early access, Mar. 4, 2021, doi: 10.1109/TCYB.2021.3054878.
- [12] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.
- [13] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Generalization of convolutional neural networks for ECG classification using generative adversarial networks," *IEEE Access*, vol. 8, pp. 35592–35605, 2020.
- [14] M. Farajzadeh-Zanjani, E. Hallaji, R. Razavi-Far, M. Saif, and M. Parvania, "Adversarial semi-supervised learning for diagnosing faults and attacks in power grids," *IEEE Trans. Smart Grid*, early access, Feb. 23, 2021, doi: 10.1109/TSG.2021.3061395.
- [15] L. A. Martínez-Insfran, F. Alconchel, P. Ramírez, P. A. Cascales-Campos, G. Carbonell, L. Barona, J. A. Pons, F. Sánchez-Bueno, R. Robles-Campos, and P. Parrilla, "Liver transplantation for fulminant hepatic failure due to heat stroke: A case report," *Transplantation Proc.*, vol. 51, no. 1, pp. 87–89, Jan. 2019.
- [16] S. Feng, N. P. Goodrich, J. L. Bragg-Gresham, D. M. Dykstra, J. D. Punch, M. A. DebRoy, S. M. Greenstein, and R. M. Merion, "Characteristics associated with liver graft failure: The concept of a donor risk index," *Amer. J. Transplantation*, vol. 6, no. 4, pp. 783–790, Apr. 2006.
- [17] A. Rana, M. A. Hardy, K. J. Halazun, D. C. Woodland, L. E. Ratner, B. Samstein, J. V. Guarrera, R. S. Brown, Jr., and J. C. Emond, "Survival outcomes following liver transplantation (SOFT) score: A novel method to predict patient survival following liver transplantation," *Amer. J. Transplantation*, vol. 8, no. 12, pp. 2537–2546, Dec. 2008.
- [18] J. B. Halldorson, R. Bakhavatsalam, O. Fix, J. D. Reyes, and J. D. Perkins, "D-MELD, a simple predictor of post liver transplant mortality for optimization of Donor/Recipient matching," *Amer. J. Transplantation*, vol. 9, no. 2, pp. 318–326, Feb. 2009.
- [19] P. Dutkowski, C. Oberkofler, K. Slankamenac, M. Puhán, E. Schadde, B. Müllhaupt, A. Geier, and P. Clavien, "Are there better guidelines for allocation in liver transplantation?: A novel score targeting justice and utility in the model for end-stage liver disease era," *Ann. Surgery*, vol. 254, no. 5, pp. 745–753, Nov. 2011.
- [20] M. Bach, A. Werner, J. Żywiec, and W. Pluskiewicz, "The study of under- and over-sampling methods' utility in analysis of highly imbalanced data on osteoporosis," *Inf. Sci.*, vol. 384, pp. 174–190, Apr. 2017.
- [21] R. Razavi-Far, M. Farajzadeh-Zanjani, and M. Saif, "An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2758–2769, Dec. 2017.
- [22] M. Farajzadeh-Zanjani, R. Razavi-Far, and M. Saif, "Efficient sampling techniques for ensemble learning and diagnosing bearing defects under class imbalanced condition," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–7, doi: 10.1109/ssci.2016.7849879.
- [23] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Inf. Sci.*, vol. 250, pp. 113–141, Nov. 2013.
- [24] A. Anand, G. Pugalenth, G. B. Fogel, and P. N. Suganthan, "An approach for classification of highly imbalanced data using weighting and under-sampling," *Amino Acids*, vol. 39, no. 5, pp. 1385–1391, Nov. 2010.
- [25] X. Tao, Q. Li, C. Ren, W. Guo, C. Li, Q. He, R. Liu, and J. Zou, "Real-value negative selection over-sampling for imbalanced data set learning," *Expert Syst. Appl.*, vol. 129, pp. 118–134, Sep. 2019.
- [26] X. Tao, Q. Li, W. Guo, C. Ren, Q. He, R. Liu, and J. Zou, "Adaptive weighted over-sampling for imbalanced datasets based on density peaks clustering with heuristic filtering," *Inf. Sci.*, vol. 519, pp. 43–73, May 2020.
- [27] Q. Cai, H. He, and H. Man, "Imbalanced evolving self-organizing learning," *Neurocomputing*, vol. 133, pp. 258–270, Jun. 2014.
- [28] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.
- [29] J. Mathew, C. K. Pang, M. Luo, and W. H. Leong, "Classification of imbalanced data by oversampling in kernel space of support vector machines," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4065–4076, Sep. 2018.
- [30] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 1322–1328.
- [31] J. Zhang and I. Mani, "KNN approach to unbalanced data distributions: A case study involving information extraction," in *Proc. ICML Workshop Learn. Imbalanced Datasets*, 2003, pp. 1–7.
- [32] S. N. N. Alifisahrin and T. Mantoro, "Data mining techniques for optimization of liver disease classification," in *Proc. Int. Conf. Adv. Comput. Sci. Appl. Technol.*, Dec. 2013, pp. 379–384.
- [33] E. M. Mirkes, T. J. Coats, J. Levesley, and A. N. Ghorban, "Handling missing data in large healthcare dataset: A case study of unknown trauma outcomes," *Comput. Biol. Med.*, vol. 75, pp. 203–216, Aug. 2016.

- [34] R. Razavi-Far, M. Farajzadeh-Zanjani, M. Saif, and S. Chakrabarti, "Correlation clustering imputation for diagnosing attacks and faults with missing power grid data," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1453–1464, Mar. 2020.
- [35] R. Razavi-Far, B. Cheng, M. Saif, and M. Ahmadi, "Similarity-learning information-fusion schemes for missing data imputation," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104805.
- [36] R. Razavi-Far, S. Chakrabarti, M. Saif, and E. Zio, "An integrated imputation-prediction scheme for prognostics of battery data with missing observations," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 1396–1404, Aug. 2013.
- [37] A. Vesin, E. Azoulay, S. Ruckly, L. Vignoud, K. Rusinová, D. Benoit, M. Soares, P. Azevedo-Maia, F. Abroug, J. Benbenishty, and J. F. Timsit, "Reporting and handling missing values in clinical studies in intensive care units," *Intensive Care Med.*, vol. 39, no. 8, pp. 1396–1404, Aug. 2013.
- [38] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, Jun. 2001.
- [39] Z. Cai, M. Heydari, and G. Lin, "Iterated local least squares microarray missing value imputation," *J. Bioinf. Comput. Biol.*, vol. 4, no. 5, pp. 935–957, Oct. 2006.
- [40] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: What is it and how does it work?" *Int. J. Methods Psychiatric Res.*, vol. 20, no. 1, pp. 40–49, Mar. 2011.
- [41] L. Gondara and K. Wang, "Recovering loss to followup information using denoising autoencoders," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 1936–1945.
- [42] D. J. Stekhoven and P. Bühlmann, "MissForest—Non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, Jan. 2012.
- [43] A. P. Robinson and J. D. Hamann, *Imputation and Interpolation*. New York, NY, USA: Springer, 2011, pp. 117–151.
- [44] W. R. Kim, T. M. Therneau, J. T. Benson, W. K. Kremers, C. B. Rosen, G. J. Gores, and E. R. Dickson, "Deaths on the liver transplant waiting list: An analysis of competing risks," *Hepatology*, vol. 43, no. 2, pp. 345–351, 2006.
- [45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2014, pp. 2672–2680.
- [46] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
- [47] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.
- [48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [49] S. Kiranyaz, O. Avcı, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107398.
- [50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–15.



**EHSAN HALLAJI** (Graduate Student Member, IEEE) received the B.Sc. degree in software engineering from Shahid Rajaei University, Tehran, Iran, in 2015, and the M.A.Sc. degree in electrical and computer engineering from the University of Windsor, Windsor, ON, Canada, in 2018, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. His current research interests include machine learning, computational intelligence, and data mining.



**ROOZBEH RAZAVI-FAR** (Senior Member, IEEE) received the B.Sc. degree in electrical and computer engineering, and the M.Sc. and Ph.D. degrees from the Amirkabir University of Technology, and the Ph.D. degree from the Politecnico di Milano. He was a Postdoctoral Fellow with the Université Libre de Bruxelles, Belgium, and an NSERC Postdoctoral Fellow and a Lecturer with the University of Windsor, Canada. He is currently with the Faculty of Engineering, University of Windsor, and serves as an Academic Advisor and a Coordinator for Master of Engineering Programs. He is also an Adjunct Professor with the Department of Electrical and Computer Engineering and a Cross-Appointed Professor with the School of Computer Science, University of Windsor. His research interests include machine learning, data mining, big data analytics, computational intelligence, cybernetics, and cybersecurity of cyber-physical systems. He is the Chapter Chair of the IEEE Computational Intelligence Society and the IEEE Systems, Man and Cybernetics Societies at IEEE Windsor Section. He serves as an Associate Editor for *Neurocomputing* journal.



**VASILE PALADE** (Senior Member, IEEE) received the Ph.D. degree from the University of Galati, Galati, Romania, in 1999. He joined Coventry University, Coventry, U.K., in 2013, after working for several years as a Lecturer with the Department of Computer Science, University of Oxford, Oxford, U.K. He has authored 200 articles in journals and conference proceedings and several books. He is currently a Professor of artificial intelligence and data science with the Centre for Data Science, Coventry University. His research interests include the area of machine learning and applications, deep learning and neural networks, various nature inspired optimization algorithms, computer vision, and natural language processing. He has delivered keynote talks and chaired international conferences on machine learning and applications. He is an associate editor for several reputed journals.



**MEHRDAD SAIF** (Senior Member, IEEE) received the B.Sc. degree in 1982, the M.Sc. degree in 1984, and the D.Eng. degree in 1987. In 1987, he joined the School of Engineering Science, Simon Fraser University. From 2002 to 2011, he was the Director of the School of Engineering Science. Since 2011, he has been the Dean of the Faculty of Engineering, University of Windsor. He has published over 350 journal articles and conference papers plus an edited book. In 1995 and 1997, he has served two terms as the Chairman of the Vancouver Section of the IEEE Control Systems Society. He is currently a member of the Editorial Board of the IEEE ACCESS and IEEE SYSTEMS JOURNAL.

...