Politecnico di Torino

Master's Degree in Data Science and Engineering

Academic Year 2021-2022

Deep Learning approaches for prediction of Hepatocellular Carcinoma Recurrence post Liver Transplantation

Supervisors

Prof. Alfredo BENSO

Prof. Gianfranco POLITANO

Dott. Quirino LAI

Candidate

Carmine DE STEFANO

Ai miei genitori,
sempre riconoscente
per avermi permesso di
raggiungere questa meta.

Summary

In recent years, several criteria have been identified for the selection of hepatocellular cancer (HCC) patients waiting for liver transplantation (LT). These criteria, like the Milan Criteria, are also the foundation of models for predicting the risk of relapse after transplantation. However, these models are severely limited in considering many variables and their non-linear interactions. This study aims to identify different deep learning models developed to improve the prediction performance of post-transplant HCC recurrence. The starting point is TRAIN-AI, based on the DeepSurv neural network: a Cox proportional hazards model. This model was developed starting from an International Cohort which will also be the main dataset of this study. Furthermore new deep learning models based on hazard rate parametrization and probability mass function (PMF) parametrization are proposed and compared, performing an appropriate hyperparameter tuning for each network. Finally a very different approach based on the DeepHit neural network is presented: this model takes into account the Competitive Risks of death.

Acknowledgements

Ringrazio i medici e i ricercatori che hanno contribuito alla raccolta dati e tutti coloro che, anche nell'attraversare un momento critico come la pandemia, hanno continuato alacremente a perseguire il miglioramento di questa realtà, con i fatti prima che con le parole.

Ringrazio il Dott. Quirino Lai per la stima dimostratami e l'opportunità di poterlo affiancare nell'importante lavoro di ricerca che è alla base di questa tesi. Ringrazio il Prof. Alfredo Benso e il Prof. Gianfranco Politano per la guida che hanno rappresentato nella stesura di questo documento.

Ringrazio Gigi per l'amicizia di cui mi onora da anni ed in particolare per il dialogo prezioso e sempre costruttivo che ha toccato i temi trattati in questa tesi.

Ringrazio la mia famiglia e i miei amici per aver sostenuto in modo diretto o indiretto i miei sforzi e le mie aspirazioni in questi anni.

Ringrazio chi mi ha donato anche solo una goccia del proprio tempo per aiutarmi a coltivare i semi della curiosità e della consapevolezza.

Table of Contents

List of Tables

List of Figures

Acronyms

HCC

Hepatocellular carcinoma

LT

Liver transplantation

LRT

Locoregional therapies

AFP

Alpha-fetoprotein

MELD

Model for End-Stage Liver Disease

LDLT

Living donor liver transplantation

HV

High volume

MC

Milano criteria

AI

Artificial Intelligence

ANN

Artificial Neural Network

PART I

Disease

CHAPTER 1
Introduction

1.1 Document guide

This thesis work is based on international research (named TRAIN-AI), in which the candidate has participated in recent years collaborating with the Organ Transplantation Unit of Sapienza University of Rome. The main purpose of this paper is to document the results achieved by using deep learning for the proposed task. It also shows possible variations that could potentially be improved. The analyses and models presented were developed using the Pycox library. In the first part there is a description of the disease and transplantation, the state of the art regarding research with similar aims and an overview of the dataset used. This is with more focus on the features chosen for the models. In the second part, all the types of networks and models used are presented with a theoretical explanation and a compilation of the most relevant results obtained. In the third part, the results of the networks are shown and compared with a reference to classical algorithms usually used for the task.

CHAPTER 2
HCC and LT

2.1 Epidemiology, aetiology, diagnosis

According to the World Health Organization in 2020, liver cancer is the sixth most common malignancy, accounting for 4.7% of all new cancer patients and the third leading cause of cancer-related deaths, accounting for 8.3% of all cancer deaths worldwide. HCC is responsible for 75–85% of all cases of primary liver cancer [1]. More than 72% of cases are estimated to occur in Asia (more than 50% in China), 10% in Europe, 7.8% in Africa, 5.1% in North America, 4.6% in Latin America, and 0.5% in Oceania [2].

Moreover, about 90% of HCCs have an underlying cause. The common aetiologies include chronic viral hepatitis B and C, alcoholic liver disease, nonalcoholic (metabolic associated) fatty liver disease, and afatoxin exposure. The incidence and aetiology of HCC vary significantly between countries and geographical regions [3]. Cirrhosis is an important risk factor for HCC, and approximately one-third of cirrhotic patients will develop

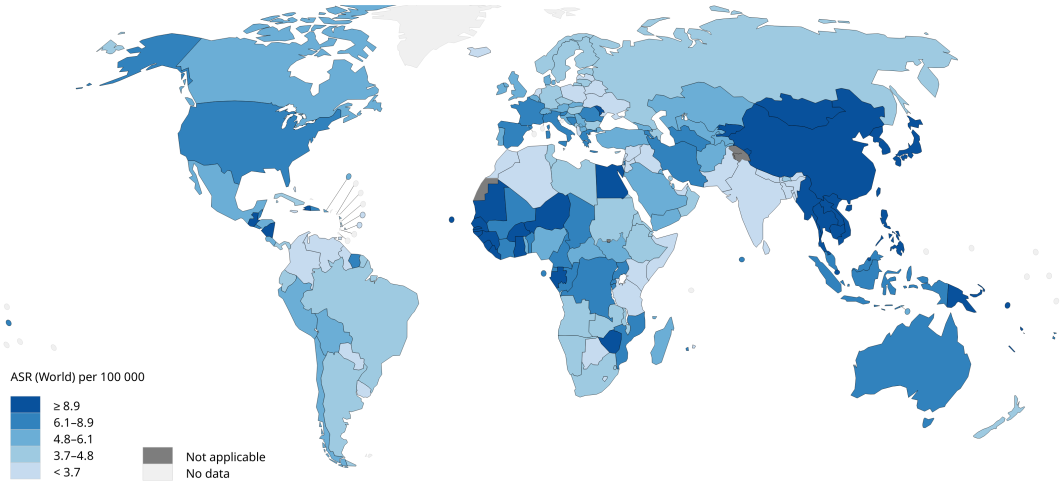


Figure 2.1: Worldwide age-standardized liver cancer incidence rates, 2020. Data source: GLOBOCAN 2020. Graph production: IARC

it in their lifetime [4][5]. Up to 80% of HCCs develop in cirrhotic patients [6]. The gold standard for diagnosing HCC is histopathological diagnosis, but liver biopsy is not routinely performed in every HCC patient due to risks of bleeding (3-4%) and needle track tumour seeding (2,7%) [4][5].

Most HCC patients are diagnosed using non-invasive radiological investigations. A liver biopsy is only recommended for patients with a liver mass with radiological features not typical of HCC, particularly in non-cirrhotic patients. The histopathological diagnosis and classification of HCC are based on the World Health Organization (WHO) classification and the International Consensus Group for Hepatocellular Neoplasia [1][7]. HCC has been shown not to be a simple homogeneous tumour and exhibits different layers of heterogeneity at the levels of aetiologies, clinical manifestations, macroscopic (radiological) appearance, histological and cytological features, genetic alterations, and clinical outcomes [8]. It is estimated that up to 30% of HCCs show distinctive morphological appearances and molecular aberrations. These have been further classified into several HCC subtypes or variants according to the WHO classification. Most cases are usually straightforward diagnoses for most general pathologists on H&E-stained sections, but diagnostic challenges may sometimes arise in cases of very well-differentiated hepatocellular lesions or small hepatic nodules (<2 cm).

Figure 2.2: Hepatocellular carcinoma. A fairly circumscribed cream-tan, solid mass with haemorrhage in a cirrhotic liver background

2.2 LT for HCC

LT is considered the most effective treatment option since it simultaneously removes the tumor and the underlying liver disease. LT was viewed as a "last chance" treatment in desperate cases of nonresectable tumors during the 80s. In 1996, the introduction of the Milan Criteria (MC) (one lesion smaller than 5 cm or up to three lesions smaller than 3 cm without extrahepatic manifestations or vascular invasion)[9] and the improvement of locoregional therapies (LRT) contributed to real progress in transplant oncology [10]. In fact, recent evidence indicates that appropriately applied LRT may increase the chance of cure, especially in cases of initially advanced tumors, outside the MC. Nowadays, MC and pre-LT LRT are widely used, resulting in excellent 5-year disease-free survival rates of 85%. There is no doubt that the MC is solid, but several groups have now extended these too restrictive MC, which denied LT access to many patients unjustifiably. Several Western and Eastern centers have proposed a prudent increase in the selection criteria, the most significant ones being San Francisco, Kyoto, Seoul, and Milan with the Metroticket concept. Resection of the liver and liver transplantation play different roles: both approaches should, however, be seen as complementary rather than competitive [11]. In cases of resectable tumors in livers with preserved function, partial hepatectomy should be used. On the other hand, liver transplantation should be used in cases of unresectable HCC or in the presence of HCC in advanced liver disease. However the high recurrence rate after hepatectomy makes LT a viable salvage therapy [12].

While tumor recurrence can reduce the usefulness of salvage LT, one of the advantages of the "resection first strategy" is the ability to obtain a precise pathological examination of the resected specimen, allowing the identification of risk factors for recurrence, such as microvascular invasion and satellite nodules [13].

CHAPTER 3
Related Works

3.1 Metroticket concept

In the previous chapters, we mentioned that several other prognostic criteria have been developed since the introduction of MC for prognostic purposes. Among the most successful is the Metroticket calculator [14], which was created by Mazzaferro and colleagues (who also developed the Milan criteria) in order to characterise more accurately the survival of patients who do not meet the Milan criteria. The results of this study have been validated in other populations, such as Chinese patients [15] and patients undergoing hepatitis B transplantation [16]. The purpose of the model was to provide a mathematical function that could predict survival after transplantation based on three covariates: size, number, and microvascular invasion. Conclusions of this study were used as a baseline for subsequent studies. From the perspective of this thesis work, the idea that we are following is to go beyond the prognostic model as a deterministic mathematical function by taking advantage of recent developments in the field of artificial intelligence.

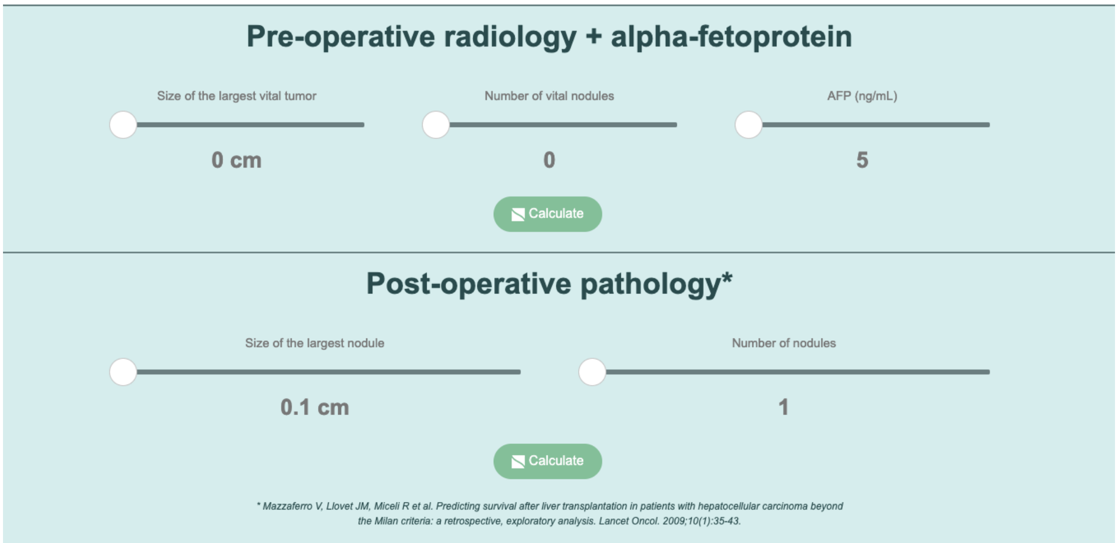


Figure 3.1: Features required by Metroticket Calculator

3.2 Machine learning and ANN

In many cases, the previously stated criteria have been combined with deterministic regression models, for example, Cox regression is widely used in survival analysis. There has been a growing number of research groups proposing to overcome the limitations presented by these approaches by using the most representative machine learning algorithms, such as Random Forest [17] or Support Vector Machine [18][19], which are suitable for classification tasks (typically the 5-year recurrence of HCC post-surgery).

Furthermore, several studies have already been conducted in the past decades that have introduced the use of neural networks for different tasks involving HCC. These include diagnosis [20], tumour progression grade [21], post resection mortality [22][23][24], and post transplant recurrence [25]. There is no doubt that these studies were pivotal in developing today's methodologies, but they were conducted on cohorts that were very limited in number, and often highly localised (it is indicative that the networks presented in such studies have one hidden layer with few perceptrons).

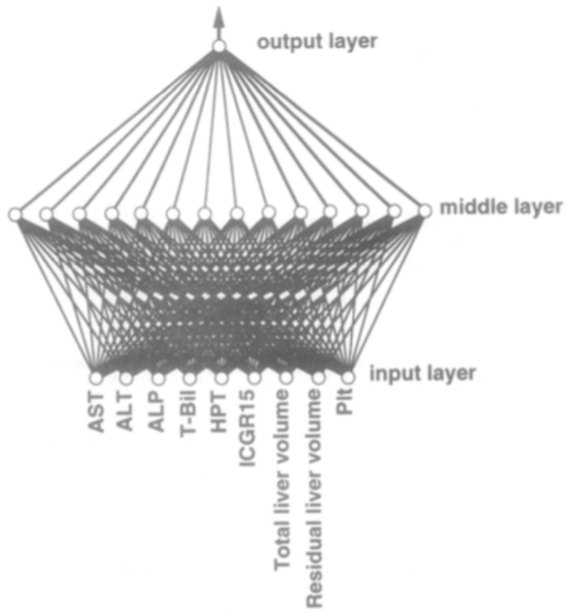


Figure 3.2: One of the first ANNs ever designed for the diagnosis of HCC [20]. It consisted of nine neurons of the input layer, 14 neurons of the middle layer and one neuron of the output.

This thesis work is the natural development of the research conducted by the candidate, which aims to develop and optimize networks capable of delivering first-order performance against the regressive task of post-LT HCC recurrence. In addition, it is fundamentally important to highlight that this has been implemented with the largest international dataset currently available, with the goal of solving the task in a global and generalized manner.

CHAPTER 4
Dataset

4.1 International Study Groups

During the course of this thesis work, the dataset that was used was the same as that used and updated by the candidate for his initial research. Several liver transplantation centers from the West and the East contributed to the development of this dataset through an extensive international effort. The collaborating centres comprised two international study groups, the EURopean HEpatocellular CAncer Liver Transplantation (EURHECALT) Study Group and the West-East Liver Transplant Study Group. A special point to highlight is that this dataset is particularly complete and integral in all its components: each record contains the data of a specific patient regarding his or her last follow-up. Consider the fact that the post-LT follow-up period for some patients can last up to 23 years in order to better highlight the importance of this collection.

RESEARCH CENTER	**LOCATION**	**PATIENTS**
Queen Mary Hospital, University of Hong Kong	Hong Kong SAR of China	260
Medanta-The Medicity	Gurgaon, India	269
Kyushu University	Fukuoka, Japan	185
Graduate School of Medicine	Kyoto, Japan	241
Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine	Kaohsiung, Taiwan	185
Medical University of Innsbruck	Innsbruck, Austria	228
Université catholique de Louvain	Brussels, Belgium	309
Merkur University of Zagreb	Zagreb, Croatia	115
Universitätsmedizin Mainz	Mainz, Germany	173
Polytechnic University of Marche	Ancona, Italy	95
University of Bologna	Bologna, Italy	472
University of Padua	Padua, Italy	437
Catholic University, Fondazione Policlinico Universitario A.Gemelli IRCCS	Rome, Italy	80
San Camillo Hospital	Rome, Italy	142
Sapienza University of Rome	Rome, Italy	205
University of Rome Tor Vergata	Rome, Italy	121
Royal Free Hospital	London, UK	153
Columbia University	New York, USA	356

Table 4.1: Dataset grouped by centers of international study groups

4.2 Hold-out strategy

The experiments in this paper were conducted using a classic hold-out strategy, which is very often used in medical research: the 4026 records were divided into three subsets that were used for training (2415), validation (805) and testing (806) of the models, respectively.

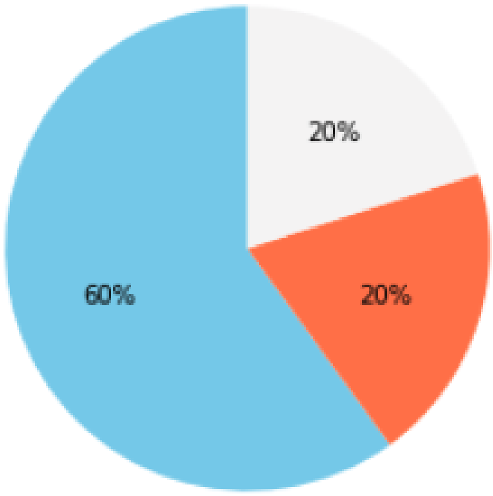


Figure 4.1: Split of dataset: 60% Training, 20% Validation, 20% Testing

By conducting a large number of experiments and past studies prior to this research [26], it was also possible to select features for the optimisation of the models, some of which were used classically for the main task, while others were introduced in an original way. The following chapter provides a detailed analysis of these.

CHAPTER 5

Features and labels

5.1 Waiting time duration

The waiting time refers to the period of time before the transplant is performed. The debate over the impact of waiting time and acceptable tumour burden on the outcomes of LT is still ongoing. An extensive multi-center study published in 2017 provides evidence of an association between very short (<6 months) or very long (>18 months) waiting times and an increased risk for HCC recurrence post-LT. A "sweet spot" of 6-18 months is therefore proposed to minimise HCC recurrence [27]. In general, experiments conducted as part of research prior to this thesis have demonstrated that the presence of this feature has an impact, albeit a limited one, on the final results.

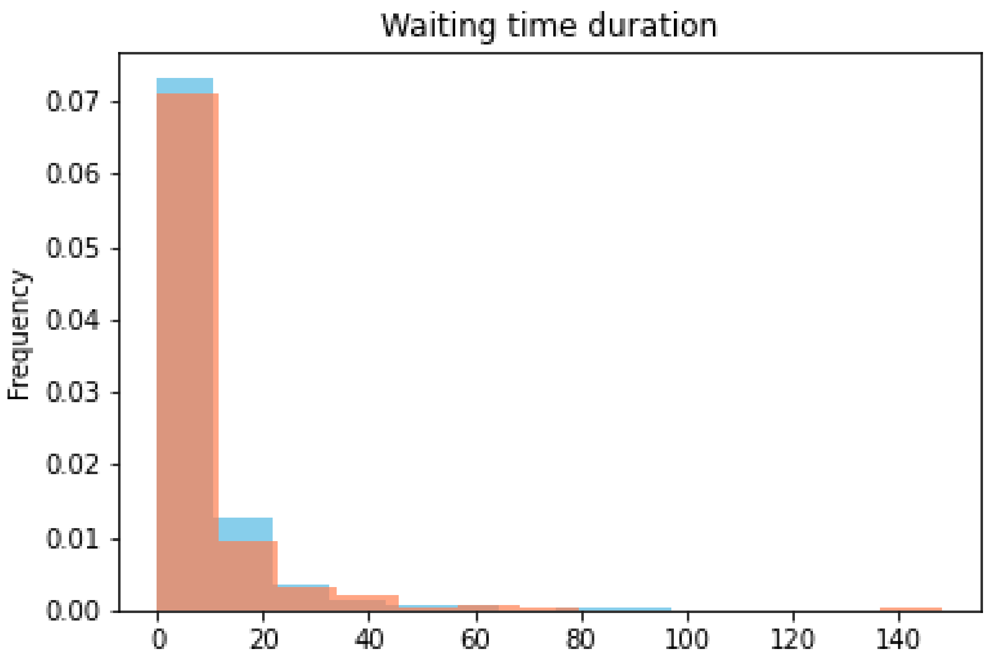


Figure 5.1: Dataset distribution: Waiting time duration in months (blue train, orange val)

5.2 Last DIAM MAX

The Last DIAM MAX is the largest diameter of the lesion that is available. There is no doubt that this is one of the most significant features for the purpose of the task. In fact, this kind of dimensional measurement is considered to be the basis for most of the criteria developed since the MC [9]. In recent times, a possible multidimensional extension of this measurement by introducing volume has been discussed [28]: this possible future implication, however, will require substantial work on aligning the datasets.

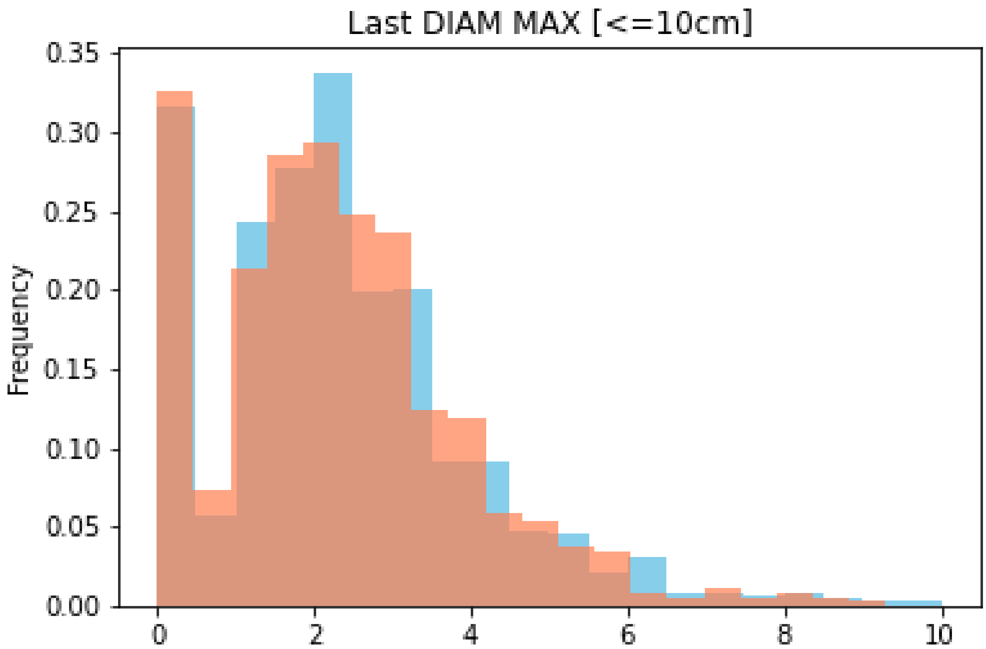


Figure 5.2: Dataset distribution: Last DIAM MAX in cm (threshold 10cm | blue train, orange val)

5.3 Last number NOD

The Last number NOD is the number of the nodules that is available. As mentioned earlier, this is one of the most important features, used in the MC [9] and the subsequent criteria.

5.4 Last log10 AFP

Last AFP stands for last available alpha-fetoprotein. It is a protein that is produced in the liver of a developing baby. The levels of AFP are usually

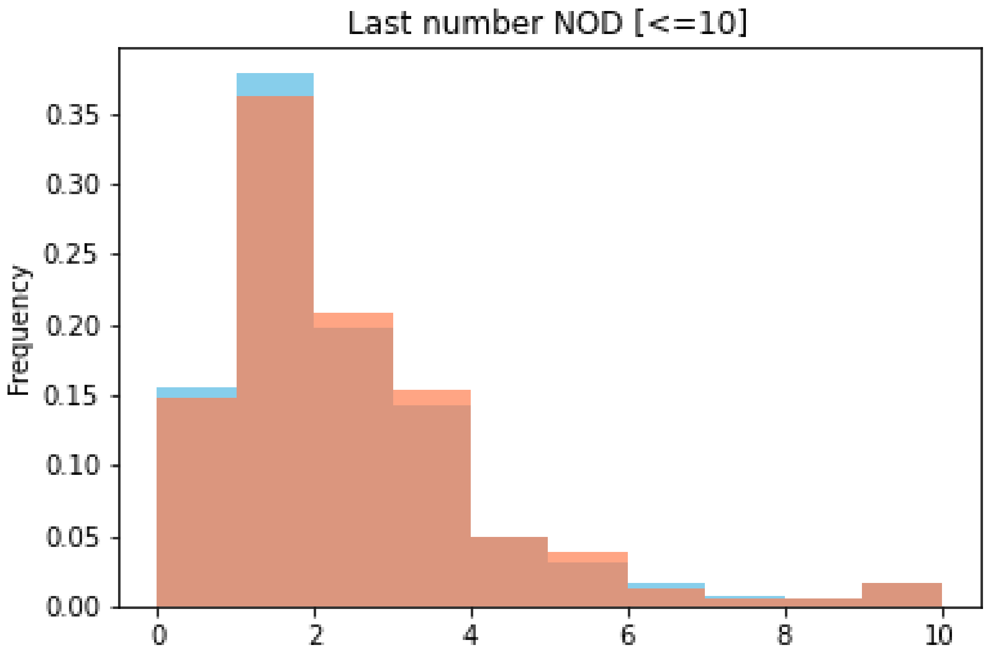


Figure 5.3: Dataset distribution: Last number NOD (threshold 10 | blue train, orange val

high when a baby is born, but they fall to very low levels by the age of one. (Healthy adults should have very low levels of this protein). AFP in serum is currently the most effective diagnostic marker for the detection of HCC. In patients with chronic liver disease, a sustained increase in the AFP serum level has been shown to be one of the risk factors for HCC, and this has been used to identify high-risk subgroups of chronic liver disease [29]. However, the exact threshold for AFP as a diagnostic criterion for HCC is controversial. [30] Furthermore, AFP may be a crucial predictor of HCC recurrence after liver transplantation [31]. The experiments in this thesis used log10 values of AFP.

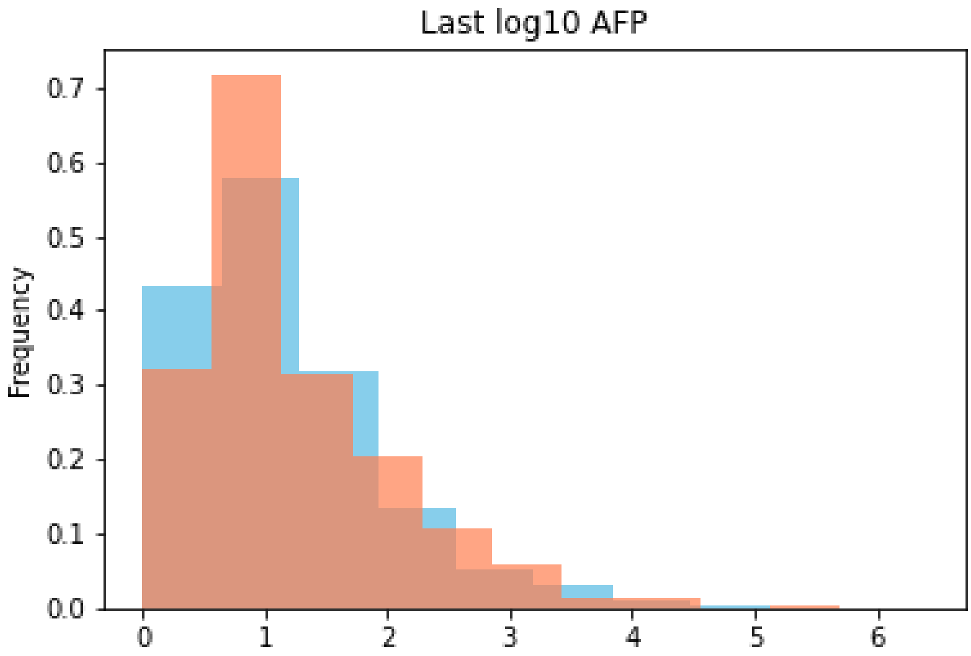


Figure 5.4: Dataset distribution: Last log10 AFP in ng/mL (blue train, orange val)

5.5 Last available MELD

The Model for End-Stage Liver Disease (MELD) score is used to estimate a patient's chances of surviving their disease over the next three months [32]. This score ranges from 6 to 40 and is based on results from several lab tests. When an organ becomes available, the higher the number, the greater the likelihood of receiving a liver from a deceased donor. The main indicators for calculating the MEDL are as follows: INR (internal normalised ratio), Creatinine, Bilirubin, Serum sodium. This score is widely accepted as a tool to prioritise organ allocation for liver transplantation.

However, it is a metric for how sick a patient is: studies [33] and experiments conducted as part of research prior to this thesis have shown that this feature can also be useful in predicting recurrence after LT.

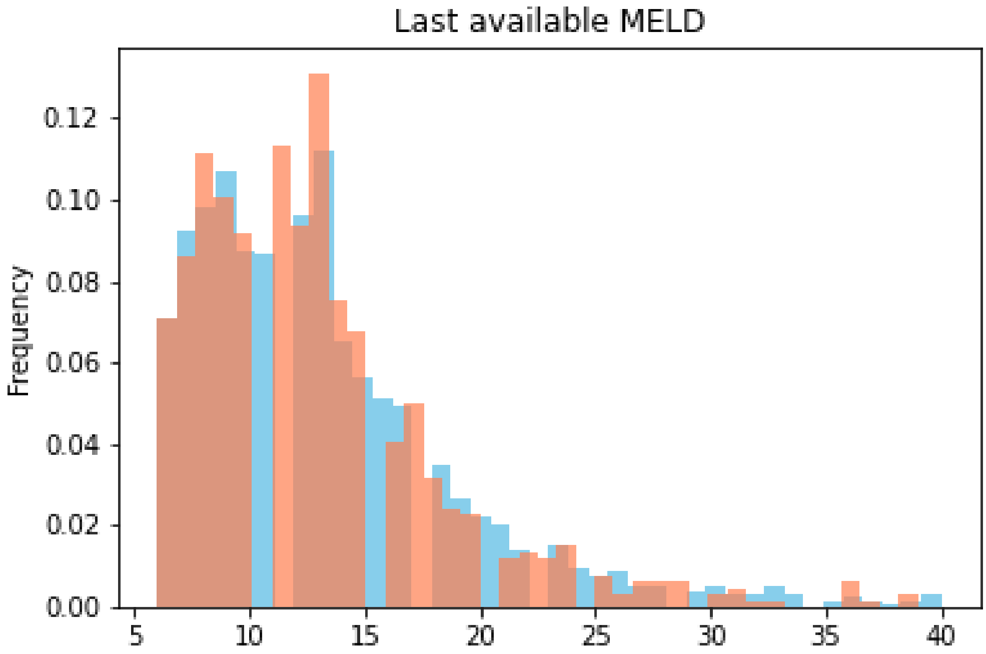


Figure 5.5: Dataset distribution: Last available MELD score (blue train, orange val)

5.6 Radiological response

The Radiological response is a feature that expresses the patient's response to treatment. LRT can play a key role in the management of hepatocellular carcinoma (HCC) as we have discussed above [34]. It has been categorically expressed as the widely known mRECIST indicator [35]. In this thesis work, the classes used are:

- Complete Response (CR)
- Partial Response (PR)
- Progressive Disease (PD)
- Stable Disease (SD) / no LRT

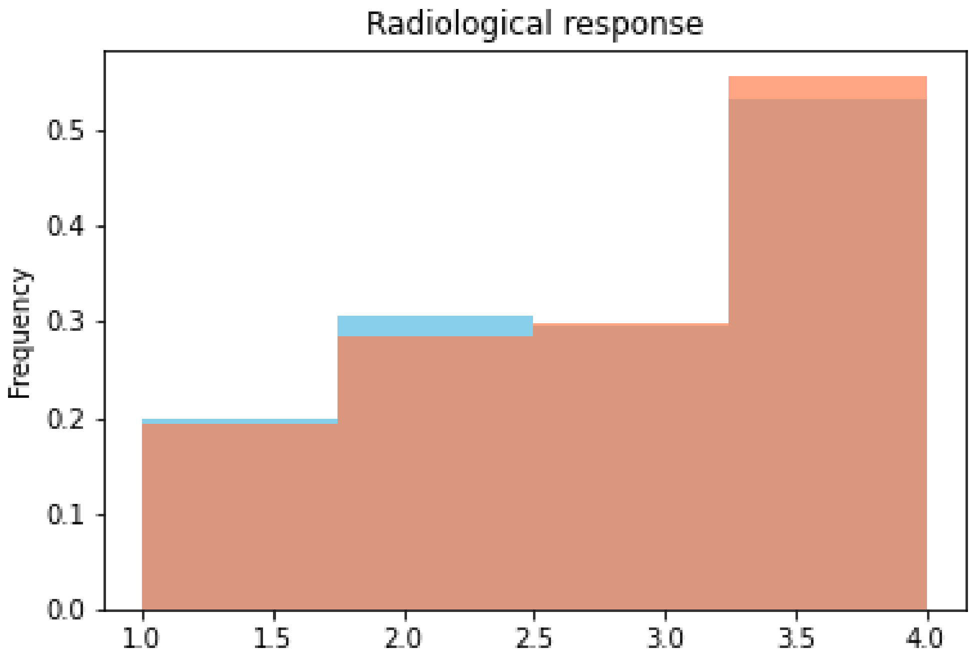


Figure 5.6: Dataset distribution: Radiological response (blue train, orange val)

5.7 Living donor LT

Living donor LT (LDLT) was initially performed almost exclusively on infants and children. In response to a critical shortage of deceased donors and an increase in waiting list mortality, adult LDLT programs were introduced several years later. The procedure is currently accepted as a therapeutic option for patients with end-stage liver disease to compensate for the shortage of organs from deceased donors [36]. Recent trends in Asian countries have shown an increased use of adult-to-adult LDLT to overcome the persistent shortage of donor organs [37] [38]. Some controversies remain in the use of LDLT, including the safety of living donation and expanding the current Milan criteria. It is generally agreed that transplant criteria should be expanded beyond the Milan criteria since the Milan criteria misses a number of patients who may benefit from LDLT [39]. This particular feature was included in this thesis since it is considered to be discriminatory for LTs and to be of fundamental importance geographically.

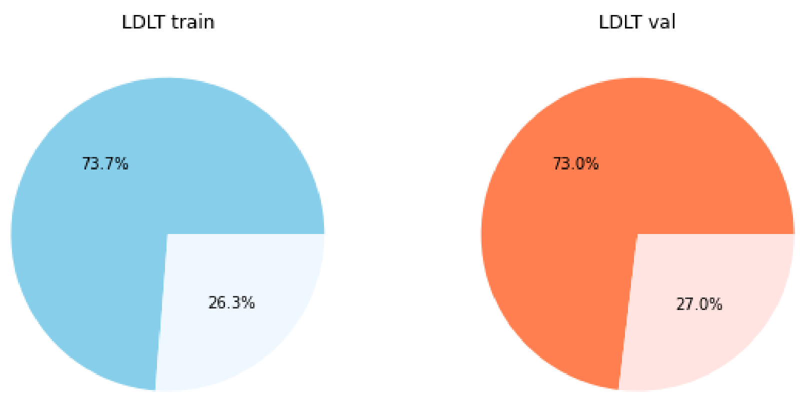


Figure 5.7: Dataset distribution: Living donor LT (light living, dark deceased | blue train, orange val)

5.8 Transplant center volume

According to several studies, the volume of transplants performed by centers is a variable of interest [40] [41]. In fact, hospital volume is associated with better overall survival, possibly due to higher treatment utilisation in high HCC volume hospitals. In this thesis work, it was decided to investigate the impact of this covariate on recurrence, so all networks were trained and optimised with and without this feature. In this dataset, it was recorded as binary: the threshold used is 70 LT/year, which is the standard in the literature.

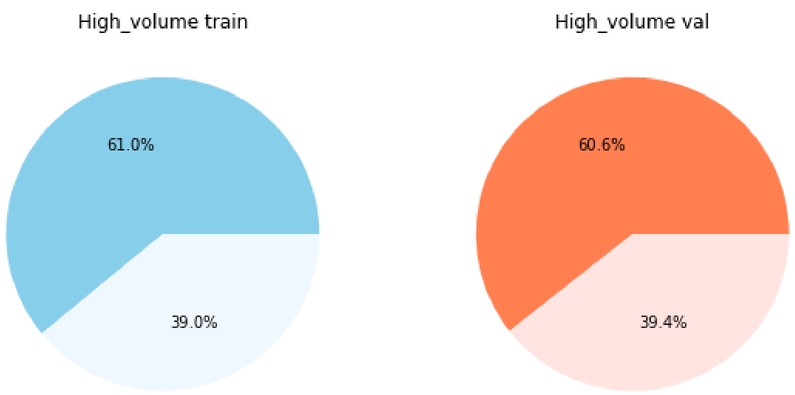


Figure 5.8: Dataset distribution: Transplant center volume in LT/year (light low, dark high | blue train, orange val)

5.9 Labels

Typically, survival analyses include two labels: the event taken into account (HCC recurrence) and the time elapsed from the LT to the event (disease free), possibly censored for death.

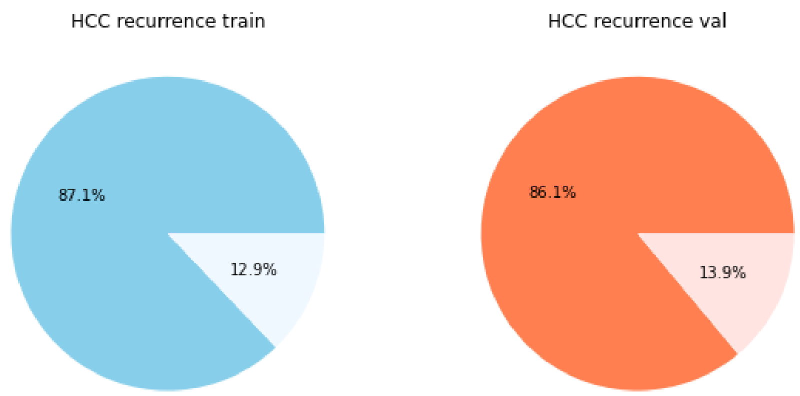


Figure 5.9: Dataset distribution: HCC recurrence (light recurrence, dark not recurrence | blue train, orange val)

In order to examine the competitive risks of death that will be investigated in the last part of this thesis, the first label will be ternary and not binary:

- Alive
- Death for HCC
- Death for other causes

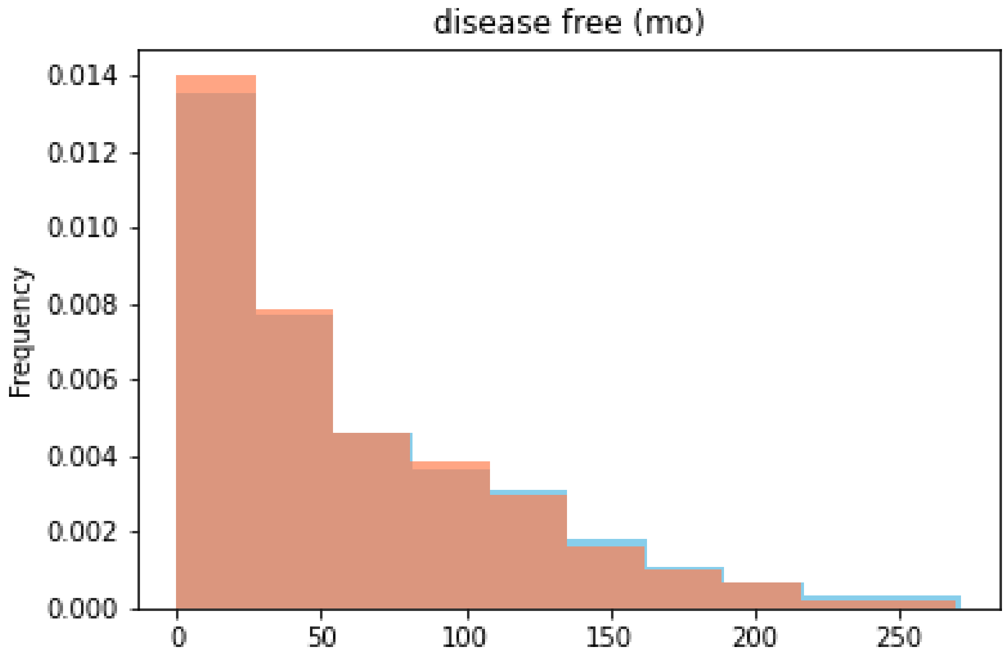


Figure 5.10: Dataset distribution: Disease free in months (blue train, orange val)

PART II

Networks

CHAPTER 6
Metric

6.1 C-index Antolini

There is a crucial aspect to developing models that can be used for outcome prediction, and that aspect is the availability of appropriate measures of predictive accuracy that can be applied to a wide range of models. Harrell's C discrimination index is one of the most common indices: an extension of the area under the ROC curve to censored survival data, which is straightforward to interpret. The original definition of C would require the prediction of individual failure times for a model with covariates with time-dependent effects and/or time-dependent features, which is not generally addressed in most clinical applications.
The metric used in the previous research and for this thesis work is a time-dependent discrimination index named C-Antolini [42], in which the whole predicted survival function is used as outcome prediction, and the ability to discriminate between subjects having different outcomes is summarized over time. This index is based on a novel definition of

concordance: a subject who develops the event should have a lower predicted probability of surviving beyond his/her survival time than any other subject who survives longer. It compares the predicted survival function of a subject who developed the event with that of subjects who developed the event before his/her survival time, and that of subjects who developed the event, or were censored, after his/her survival time. It is important to note that censored subjects are included in comparisons with subjects who developed the event before their observed times. In addition, the Kaplan-Meier (KM) estimator is also used to treat these subjects in evaluation step. KM [43] is the most widely used survival model in the statistical and medical literature, which has the advantage of learning very flexible survival curves, but the disadvantage of not taking into account patients' covariates. Therefore, it is useful at the population level, but not at the individual level, requiring more complex techniques such as regressors and deep learning, as shown in this thesis.

CHAPTER 7
DeepSurv

7.1 Background

Prior research (TRAIN-AI) has been based exclusively on the DeepSurv model [44]: a deep neural network with Cox proportional hazards and a state-of-the-art survival method that takes into account the interaction between a patient's features and treatment efficacy to provide personalised treatment recommendations.

The term survival refers to a specific event (such as the recurrence of illness or death). The survival function and the hazard function are the two fundamental functions of survival analysis. Survival function $S(t) = Pr(T > t)$ indicates the probability that an individual has "survived" beyond time t. The hazard function $\lambda(t)$ is defined as:

$$\lambda(t) = \lim_{\delta \to 0} \frac{\Pr(t \leq T < t + \delta | T \geq t)}{\delta} \tag{7.1}$$

The hazard function is the probability that an individual will not survive an additional infinitesimal amount of time δ, assuming they have already survived to time t. Thus, a greater hazard signifies a greater risk.

In the model, the hazard function is composed of two non-negative functions: a baseline hazard function, $\lambda_0(t)$, and a risk score, $r(x) = e^{h(x)}$, defined as the effect of an individual's observed covariates on the baseline hazard function. We denote $h(x)$ as the log-risk function. The hazard function is assumed to have the following form:

$$\lambda(t|x) = \lambda_0(t) \cdot e^{h(x)} \tag{7.2}$$

In the linear case, the proportional hazards model (CPH) is a proportional hazards model that estimates the log-risk function, by linear function $\widehat{h}_\beta(x) = \beta^T x$. In the non-linear case of neural networks [45], the linear combination of features $\widehat{h}_\beta(x)$ is replaced with the output of the network $\widehat{h}_\theta(x)$. In fact, DeepSurv is a deep feed-forward neural network that predicts a patient's hazard rate based on the weights of the network θ. The input to the network is a patient's baseline data $\mathbf{x}$. The hidden layers of the network consist of a fully-connected layer of nodes, followed by a dropout layer. The output of the network $\widehat{h}_\theta(x)$ is a single node with a linear activation which estimates the log-risk function in the Cox model. The network is trained by setting the objective function to be the average negative log partial likelihood with regularization:

$$L(\theta) = -\frac{1}{N_{E=1}} \sum_{i:E_i=1} \left(\widehat{h}_\theta(\mathbf{x}_i) - \log \sum_{j \in \Re(T_i)} e^{\widehat{h}_\theta(\mathbf{x}_j)} \right) + \lambda \cdot \|\theta\|_2^2 \tag{7.3}$$

where $N_{E=1}$ is the number of patients with an observable event and

λ is the l_2 regularization parameter. Gradient descent optimization is then used to find the weights of the network that minimize the previous equation.

Modern deep learning techniques are used to optimize the training of the network. These include standardizing the input, Rectified Linear Units (ReLU) as the activation function, Adaptive Moment Estimation (Adam) for gradient descent, and learning rate scheduling.

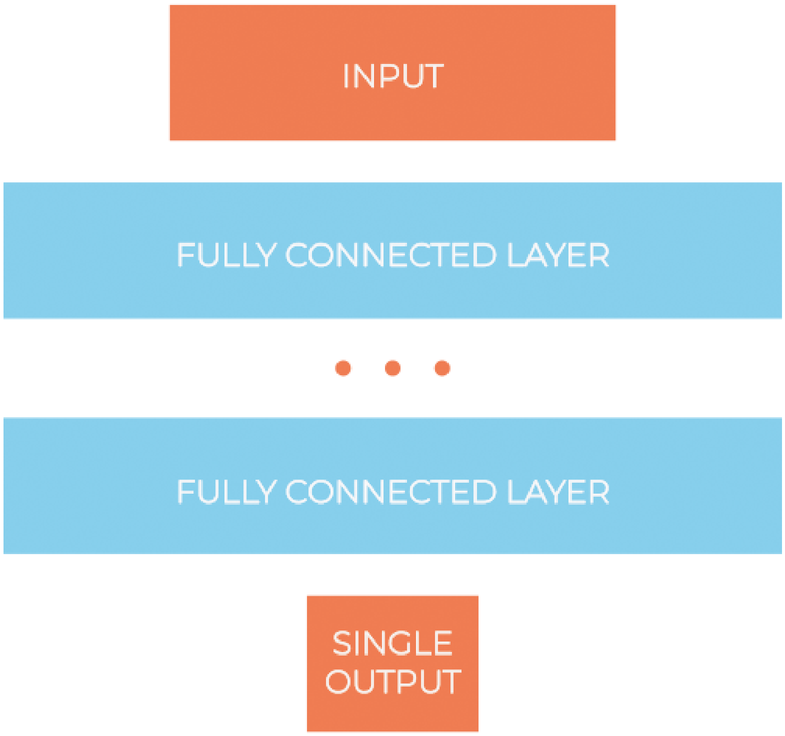


Figure 7.1: DeepSurv architecture

7.2 Method

In this configuration, the number and density of the layers is primarily determined by the need to avoid problems of too much simplicity in predicting Val and Test (which would result in a loss on the Val less than that on the Train, which is exactly what should be avoided). Nevertheless, an appropriate number of epochs was also chosen (as well as the epoch at which the LR was rescaled) in order to avoid overfitting while still achieving the highest possible metrics.

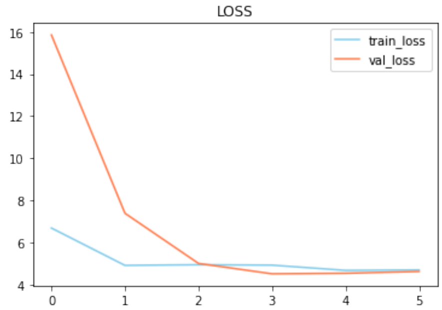


Figure 7.2: DeepSurv1

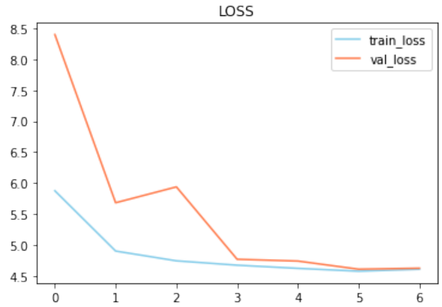


Figure 7.3: DeepSurv2

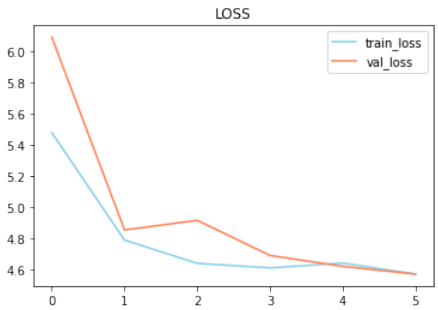


Figure 7.4: DeepSurv3 (+HV)

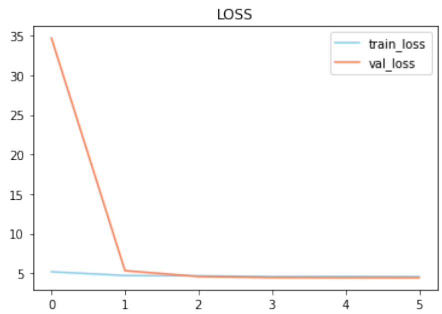


Figure 7.5: DeepSurv4 (+HV)

	DeepSurv1	**DeepSurv2**	**DeepSurv3**	**DeepSurv4**
LAYERS	[512]*4	[512]*5	[512]*4 [256]	[512]*4 [256] [128]
DROPOUT	0.1	0.1	0.1	0.1
LR	ep0 : 5e-3	ep0 : 6e-3 ep5 : 1e-4	ep0 : 8e-3 ep3 : 1e-4	ep0 : 9e-3 ep3 : 1e-4
C train	0.77	0.78	0.78	0.77
C val	0.75	0.74	0.75	0.76
C test	0.75	0.74	0.75	0.76

Table 7.1: DeepSurv models

CHAPTER 8
PC Hazard

8.1 Background

The approach proposed in this chapter is structurally and conceptually similar to the previous one, but with one significant difference: it is a continuous-time approach that assumes that the continuous-time hazard rate is piecewise constant [46]. The temporal discretisation methods used in this and subsequent sections do not change the nature of the task since the dataset used has monthly granularity for temporal features and labels.

First of all consider a partition of the time scale τ and $k(t)$ denoting interval index of time t such that $t \in (\tau_{k(t)-1}, \tau_{k(t)}]$. Taking the assumption that the hazard is constant within each interval, we can express the hazard as a step function:

$$h(t) = \eta_{k(t)} \tag{8.1}$$

for a set of non-negative constants $\{\eta_1, ..., \eta_m\}$. Therefore, the loss used for this network is the mean negative log-likelihood:

$$L = -\frac{1}{n}\sum_{i=1}^{n}\left(d_i \log \widetilde{\eta}_{k(t)}(\mathbf{x}_i) - \widetilde{\eta}_{k(t_i)}(\mathbf{x}_i)\rho(t_i) - \sum_{j=1}^{k(t_i)-1}\widetilde{\eta}_j(\mathbf{x}_i)\right) \tag{8.2}$$

with $d_i \in \{0,1\}$, $\widetilde{\eta}_j = \eta_j \Delta\tau_k$, $\rho(t) = \frac{t-\tau_{k(t)-1}}{\Delta\tau_{k(t)}}$.

To follow the complete development of the formulae, see [46].

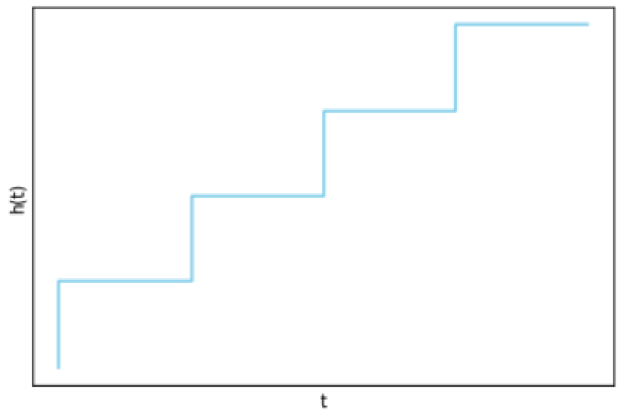


Figure 8.1: PC Hazard conceptual representation

8.2 Method

The most promising results from the numerous trials conducted are presented below. In comparison to its predecessor, the network does not improve in terms of performance, but it allows for comparable metrics with a 'lighter' structure in terms of layers. The loss is generally more stable.

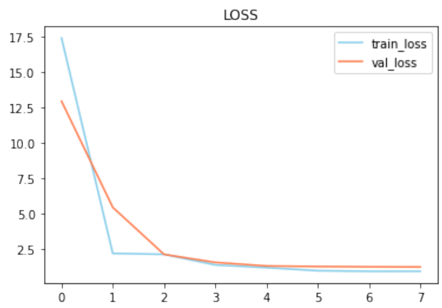


Figure 8.2: PCHazard1

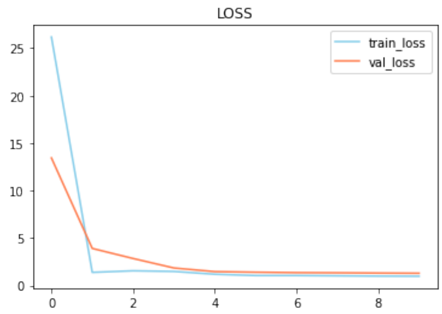


Figure 8.3: PCHazard2

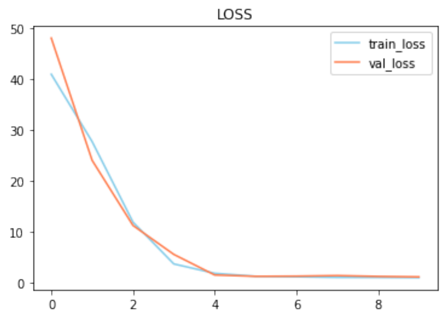


Figure 8.4: PCHazard3 (+HV)

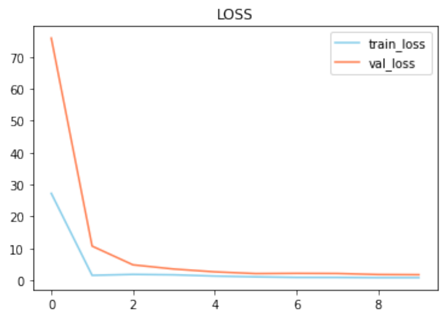


Figure 8.5: PCHazard4 (+HV)

	PCHazard1	**PCHazard2**	**PCHazard3**	**PCHazard4**
LAYERS	[64]*2	[128]*7	[64]*4	[128]*6
DROPOUT	0.1	0.2	0.1	0.1
LR	ep0 : 1e-1 ep5 : 1e-3	ep0 : 4e-2 ep5 : 5e-3	ep0 : 1e-2	ep0 : 4e-2
C train	0.72	0.73	0.74	0.75
C val	0.72	0.72	0.74	0.75
C test	0.72	0.73	0.73	0.75

Table 8.1: PCHazard models

CHAPTER 9
Neural-MTLR

9.1 Background

The likelihood for discrete-time survival data may be parameterised by the discrete hazard rate but also by a probability mass function (PMF). This type of approach requires fewer assumptions than the previous ones and has been developed in several ways. This approach is used in this chapter through the Multi-task Logistic Regression (MTLR) method applied to a neural network.

The original MTLR [47] is a generalization of the binomial log-likelihood to jointly model the sequence of binary labels. In fact $Y = (y_1, ..., y_m)$ is a sequence with zeros for every time τ_j up to the event time, followed by one's, e.g., $(0, ..., 0, 1, ...1)$.

$$\Pr(Y = (y_1, ..., y_m)|\mathbf{x}) = \frac{\exp[\sum_{k=1}^{m} y_k \psi_k(\mathbf{x})]}{1 + \sum_{k=1}^{m} \exp[\sum_{l=k}^{m} \psi_l(\mathbf{x})]} \tag{9.1}$$

But one problem remains: the model is still linear at its core, with $\psi k(\mathbf{x}) =$

$\mathbf{x}^T\beta_k$, so it cannot properly model nonlinear dependencies in the dataset. Therefore, this model has been extended to the N-MTLR [48]: in particular the parameters of $\psi k(\mathbf{x})$ are found by minimizing the negative log-likelihood:

$$L = -\frac{1}{n}\sum_{i=1}^{n}(d_i \log[f(t_i|\mathbf{x_i})] + (1-d_i)\log[S(t_i|\mathbf{x_i})] \tag{9.2}$$

with $f(\tau_j) = \Pr(T = \tau_j)$ as PMF, $S(\tau_j) = \sum_{k>j} f(\tau_j)$ as survival function.

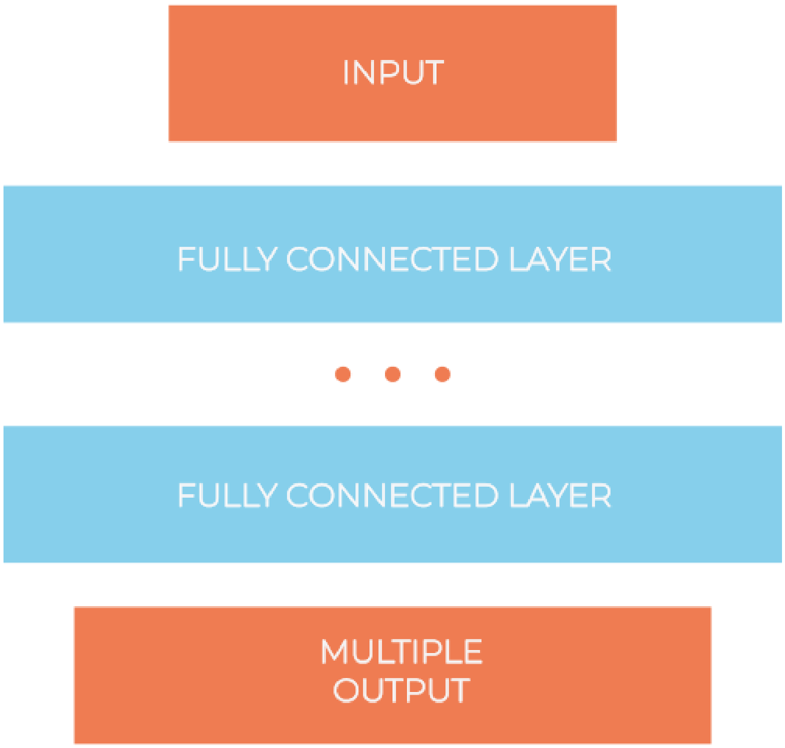


Figure 9.1: Neural-MTLR architecture

9.2 Method

The minimum granularity on the timescale in the dataset (monthly) was used for necessary discretization, and it is notable that the network manages to achieve comparable performance to the previous ones, but with a "heavier and deeper" structure at the layer level that requires more training epochs. Nevertheless, it may be a viable alternative if a more pronounced discretisation of the time span is required, e.g. with annual granularity (although this obviously worsens the metrics because the information content is synthesised).

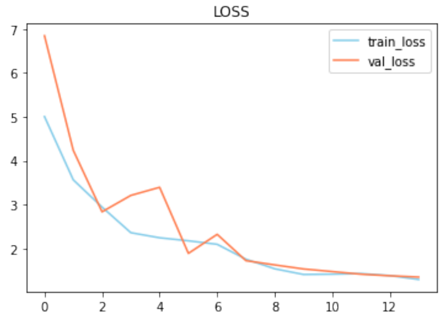


Figure 9.2: N-MTLR1

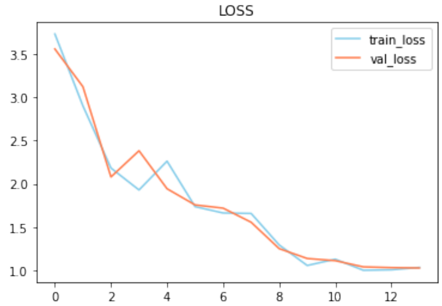


Figure 9.3: N-MTLR2

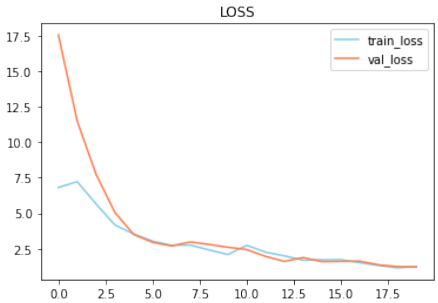


Figure 9.4: N-MTLR3 (+HV)

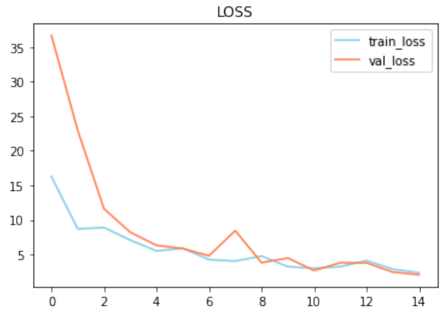


Figure 9.5: N-MTLR4 (+HV)

	N-MTLR1	**N-MTLR2**	**N-MTLR3**	**N-MTLR4**
LAYERS	[128]*5	[128]*5	[256]*10	[512]*6
DROPOUT	0.1	0.1	0.2	0.1
LR	ep0 : 1e-2 ep8 : 1e-4	ep0 : 8e-3 ep12 : 1e-4	ep0 : 9e-3 ep18 : 5e-4	ep0 : 1e-2
C train	0.74	0.75	0.77	0.76
C val	0.72	0.74	0.74	0.75
C test	0.73	0.73	0.74	0.75

Table 9.1: N-MTLR models

CHAPTER 10
DeepHit - Competing Risks

10.1 Background

Researchers have increasingly focused on developing predictive tools that can distinguish between specific risks, especially in the case of death. A survival analysis with competing risks is a challenging problem. This is made all the more relevant because the choice of treatment must take into account these competing risks. Furthermore, right-censoring of data is extremely common in the medical setting: patients are frequently lost to follow-up (often for unknown reasons). One of the proposals in this area is the Metroticket 2.0 [49], an evolution of the model presented in one of the previous chapters. It is a combination of classical regression models focusing on the different risks to be considered. This type of task can also be handled by deep learning: in this last part of the thesis work, a deep neural network was implemented and used that directly learns the distribution of first hit times, hence the name DeepHit [50]. It can be considered one of the parameterisation models of the PMF (like the N-MTLR in the previous chapter). Therefore, it is possible to avoid the strong assumptions of models derived from Cox regression on the relationship between the covariates and the parameters of this process. This is the case in one of the most popular models for competitive risks, Fine and Gray [51]. In this case, the goal remains to calculate Cumulative Incidence Functions (CIF), which estimate the marginal probability for each competing event. Marginal probability refers to the probability of the individuals who developed the event of interest, regardless of whether they were censored or unsuccessful for other competing events.

DeepHit uses a multi-task network architecture that consists of a shared sub-network and a family of cause-specific sub-networks. This architecture differs slightly from that of conventional multi-task networks. In fact, it is implemented with a single softmax layer as the output layer of DeepHit in order to ensure that the network learns the joint distribution of K competing events not the marginal distributions of each event. A loss function is used to train by exploiting both survival times and relative risks. This loss function is the sum of two terms $L_{Total} = L_1 + L_2$.

L_1 is the log-likelihood of the joint distribution of the first hitting time and event: it drives DeepHit to learn the general representation for the joint distribution of the first hitting time and event.

$$\begin{aligned} L_1 = -\sum_{i=1}^{N} \Bigg[& \nVdash(k^{(i)} \neq \emptyset) \cdot \log(y^{(i)}_{k^{(i)},s^{(i)}}) \\ & + \nVdash(k^{(i)} \neq \emptyset) \cdot \log\left(1 - \sum_{k=1}^{K} \widehat{F}_k(s^{(i)}|\mathbf{x}^{(i)})\right) \Bigg] \end{aligned} \tag{10.1}$$

Each patient history is prepresented by a triple $(\mathbf{x}, s, k)$ where $x \in X$ is a D-dimensional vector of covariates, $s \in T$ is the time at which the event or censoring occurred, and $k \in K$ is the event or censoring that occurred at time s. Right-censoring id indicated by $\emptyset$. The output of the softmax layer is a probability distribution y.

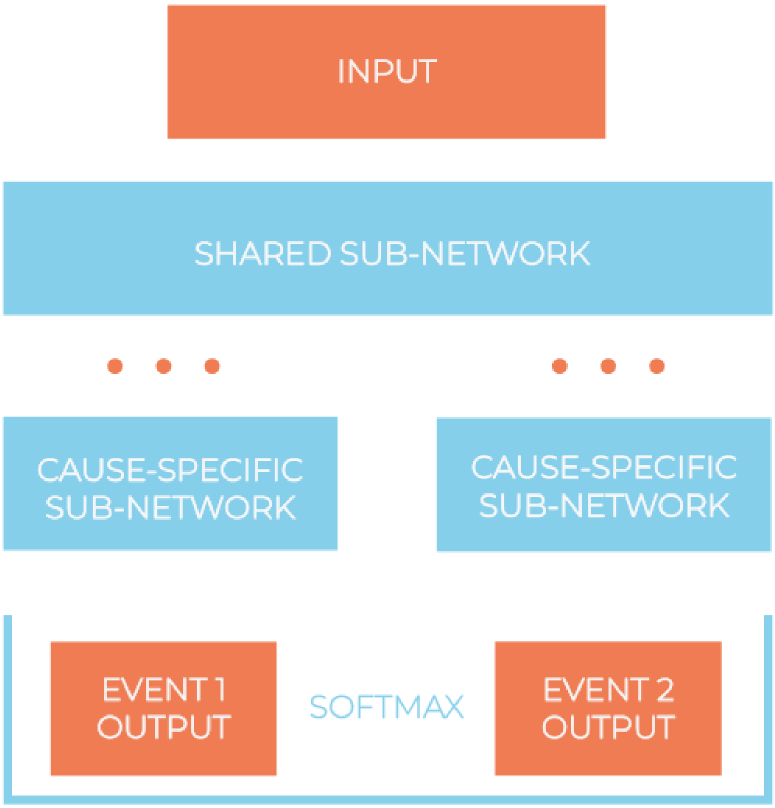


Figure 10.1: DeepHit architecture

L_2 incorporates estimated CIFs calculated at different times (i.e. the time at which an event actually occurs) in order to fine-tune the network to each cause-specific estimated CIF expressed as a function $\hat{F}_k$.

$$L_2 = \sum_{k=1}^{K} \alpha_k \cdot \sum_{i \neq k} A_{k,i,j} \cdot \eta\Big(\widehat{F}_k(s^{(i)}|\mathbf{x}^{(i)}), \widehat{F}_k(s^{(i)}|\mathbf{x}^{(j)})\Big) \tag{10.2}$$

$A_{k,i,j} = \nVdash(k^{(i)} = k, s^{(i)} < s^{(j)})$ in an indicator function of pairs (i, j) who experience risk k at different time. The coefficients α_k are chosen to trade off ranking losses of the k-th competing event, and $\eta(x, y)$ is a convex loss function. To follow the complete development of the formulae, see [50].

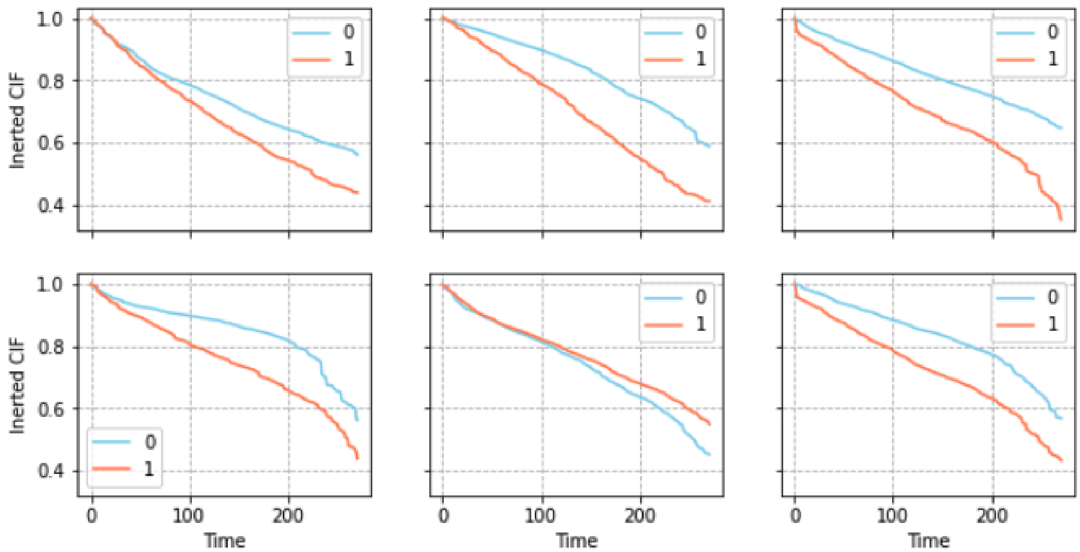


Figure 10.2: Inverted CIFs of six random patients (blue HCCdeath, orange OTHERdeath)

10.2 Method

Below are the best results, however it is important to note that these results should not be compared with the previous networks as a different objective is announced: the specific risk of death from HCC instead of recurrence risk. The model calculates the so-called CIFs for competitive risks and then calculates the metrics using these inverted curves (CIFs increase with risk).

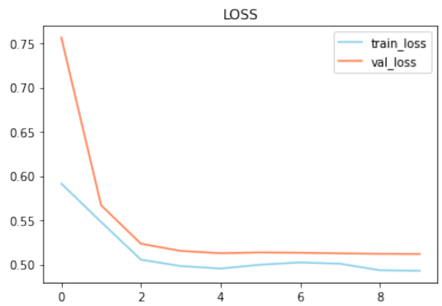


Figure 10.3: DeepHit1

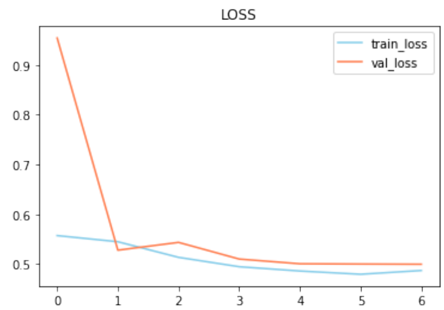


Figure 10.4: DeepHit2

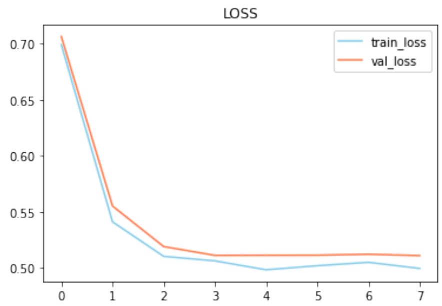


Figure 10.5: DeepHit3 (+HV)

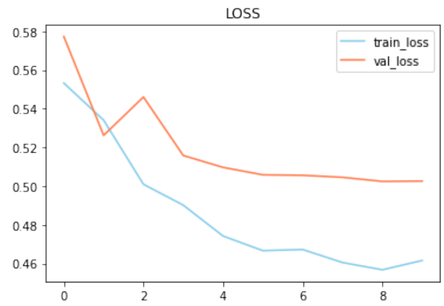


Figure 10.6: DeepHit4 (+HV)

Another aspect to note about this network is that it is trained using the

AdamWR cyclic optimiser with following regularisation policies:
$decoupled\_weight\_decay = 0.01$ and $cycle\_eta\_multiplier = 0.6$.

	DeepHit1	**DeepHit2**	**DeepHit3**	**DeepHit4**
LAYERS shared	[512]*5	[256]*7	[512]*10	[256]*7
LAYERS individual_risk	[256]*4	[128]*3	[256]*4	[128]*3
DROPOUT	0.1	0.1	0.1	0.1
LR	ep0 : 1e-2 ep2 : 1e-4	ep0 : 1e-2 ep4 : 1e-4	ep0 : 1e-2 ep3 : 1e-4	ep0 : 1e-2 ep4 : 1e-4
C train HCC death	0.75	0.76	0.74	0.79
C train other death	0.61	0.63	0.56	0.66
C val HCC death	0.73	0.73	0.73	0.76
C val other death	0.58	0.57	0.54	0.57
C test HCC death	0.75	0.75	0.73	0.78
C test other death	0.57	0.58	0.51	0.56

Table 10.1: DeepHit models

PART III

Results

CHAPTER 11
Work analysis

11.1 Comparison with classical criteria

The last part of the thesis compares the metrics obtained with the most effective models presented above (with the exception of DeepHit, which was developed for a different purpose) and those obtained from the classical algorithms underlying the commonly applied regressors. A comparison was conducted in the Test set, and it is pertinent to emphasize that the metrics were calculated after making a cut based on the criteria scores according to the common practice of considering the fifth year post-LT. To ensure stability and reliability, a 1000 repetition bootstrap was conducted. In accordance with this procedure, the 95% confidence intervals were calculated using the established percentile method. The following two tables show the results for the models that used seven features and those that also used the High Volume feature: as can be seen here, a slight improvement in performance has been made. All deep learning models perform significantly better than classical algorithms (of which Metroticket is the most efficient). However, the DeepSurv models seem to be the best, which shows that for a relatively large amount of data, discretisation methods do not pay off, and continuous-time methods based on parameterisation of the hazard rate are preferable.

	C-Antolini	**Lower**	**Upper**
Milan	0.63	0.58	0.68
San Francisco	0.60	0.56	0.65
Upto7	0.59	0.56	0.65
Metroticket	0.66	0.61	0.71
HALTHCC	0.52	0.50	0.55
AFPFrench	0.64	0.58	0.68
ASAN	0.60	0.56	0.64
Kyoto	0.58	0.54	0.62
DeepSurv1	0.75	0.69	0.80
PCHazard2	0.73	0.67	0.78
MTLR2	0.73	0.68	0.78

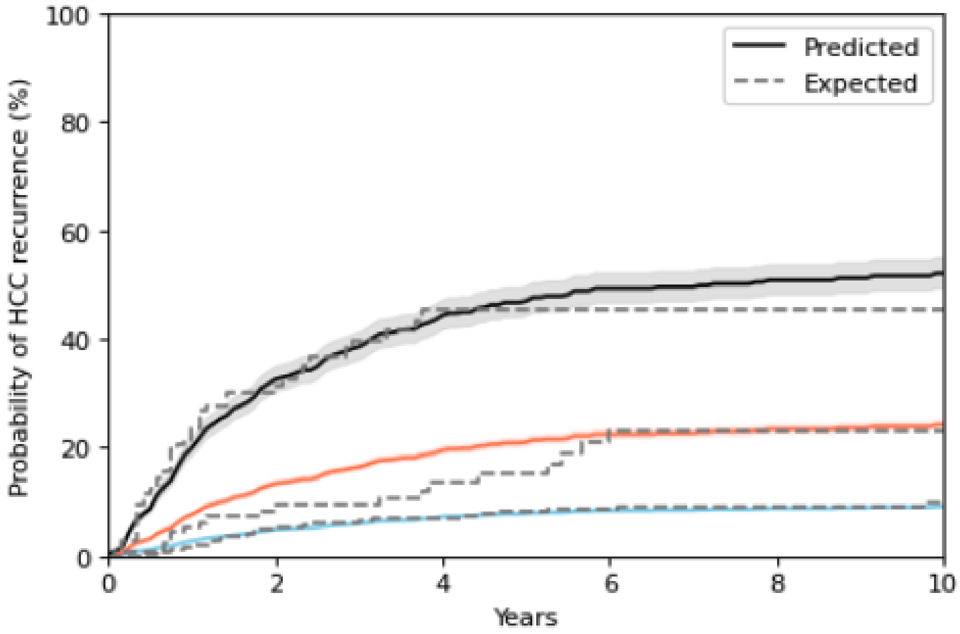
Table 11.1: Test comparison (CI 95%) without High Volume feature

	C-Antolini	Lower	Upper
Milan	0.63	0.58	0.68
San Francisco	0.60	0.56	0.66
Upto7	0.59	0.55	0.64
Metroticket	0.66	0.61	0.71
HALTHCC	0.52	0.50	0.55
AFPFrench	0.64	0.59	0.69
ASAN	0.60	0.56	0.64
Kyoto	0.58	0.54	0.62
DeepSurv1	0.76	0.70	0.81
PCHazard2	0.75	0.70	0.80
MTLR2	0.75	0.68	0.80

Table 11.2: Test comparison (CI 95%)
with High Volume feature

11.2 Practical interpretability

For predictive computer models, it is crucial to understand how the results can be interpreted and applied in practice. In this specific task, each predictive model was compared with reality for three risk classes (thresholds: 0.70 and 0.85). To obtain the predicted curve, the Kaplan-Meier algorithm is used on the Test set. The mean probability curve of the predictions was plotted over time (with 95% confidence intervals). Curves are calculated over the original time period, but trimmed to 10 years (not very significant later).



	1	2	3	4	5	6	7	8	9	10
High risk	111	76	55	39	29	24	22	18	17	12
Moderate risk	127	105	84	71	61	48	38	32	25	20
Low risk	568	471	388	341	304	251	227	179	147	112

Figure 11.1: Risk ranges using DeepSurv4

CHAPTER 12

Future developments

12.1 Goals and perspectives

This thesis shows that deep learning has many application possibilities in the field of survival analysis. The proposed approaches make the most of the large amount of data collected and make this type of cooperation between databases vital for the future.

Another important aspect of this last part is the construction and development of simple and immediate tools that will allow people without advanced computer or AI skills to use these models. In this sense, taking a cue from the Metroticket, a web application capable of serving doctors and researchers who would like to use the best model presented here was developed during the thesis phase.

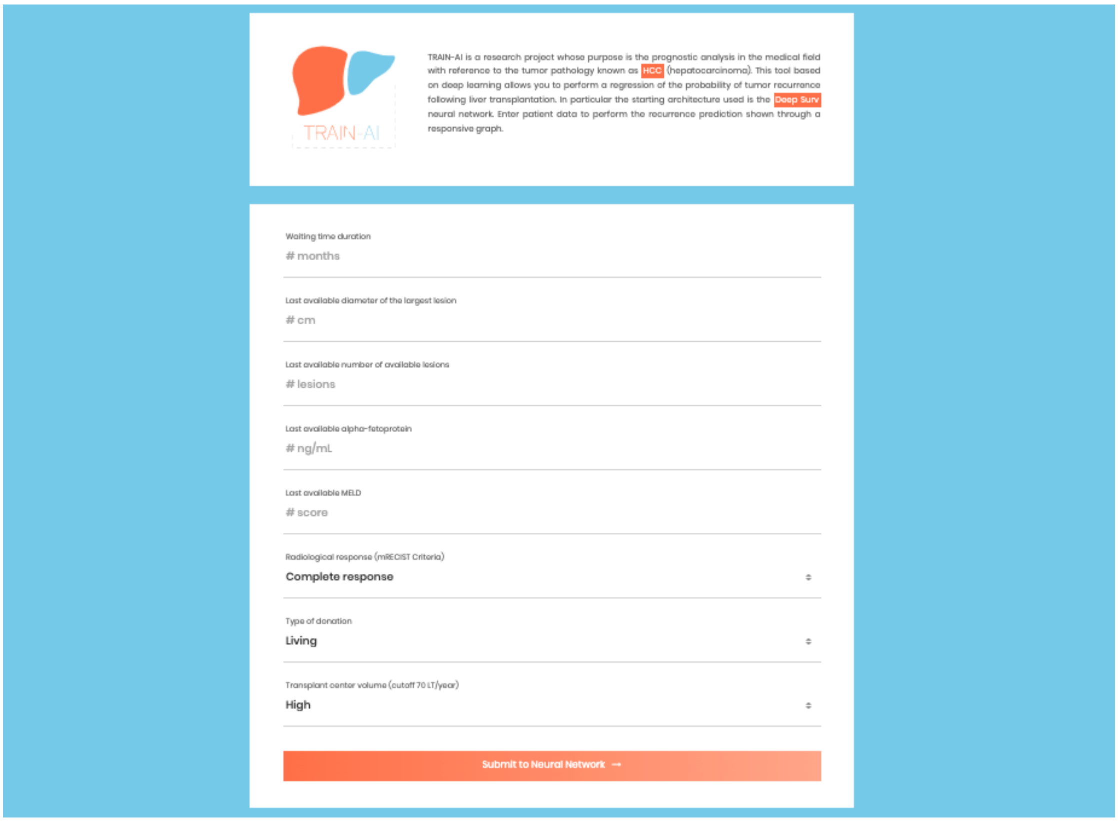


Figure 12.1: Features required by TRAIN-AI Calculator

Lastly, it is important to consider the future impact that the approaches presented in similar tasks may have, which is even more relevant than the ones analyzed: if in fact the MC has been developed to improve the criteria of the transplant list, but it is still widely used today also for the prognostic of post-LT recurrence, it will be possible to perform the "reverse path" for deep learning models. By combining different features and optimizing them appropriately, it will be possible to develop models that will significantly improve the functioning of transplant lists in the future.

Bibliography

[1] M. S. Torbenson, I. O. L. Ng, Y. N. Park, M. Roncalli, and M. Sakamato. «WHO Classification of Tumours. Digestive System Tumours». In: *Lyon, France: International Agency for Research on Cancer* (2019), pp. 229–39 (cit. on pp. 3, 4).

[2] A. G. Singal, P. Lampertico, and P. Nahon. «Epidemiology and surveillance for hepatocellular carcinoma: New trends». In: *Journal of hepatology* 72 (2 2020), pp. 250–261. DOI: https://doi.org/10.1016/j.jhep.2019.08.025. URL: https://pubmed.ncbi.nlm.nih.gov/31954490/ (cit. on p. 3).

[3] J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray. «Cancer statistics for the year 2020: An overview.» In: *International journal of cancer* (2021). DOI: https://doi.org/10.1002/ijc.33588. URL: https://pubmed.ncbi.nlm.nih.gov/33818764/ (cit. on p. 3).

[4] European Association for the Study of the Liver. «EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma». In: *Journal of hepatology* 69 (1 2018), pp. 182–236. DOI: https://doi.org/10.1016/j.jhep.2018.03.019. URL: https://pubmed.ncbi.nlm.nih.gov/29628281/ (cit. on p. 4).

[5] J. K. Heimbach, L. M. Kulik, R. S. Finn, C. B. Sirlin, M. M. Abecassis, L. R. Roberts, A. X. Zhu, M. H. Murad, and J. A. Marrero. «AASLD guidelines for the treatment of hepatocellular carcinoma». In: *Hepatology (Baltimore, Md.)* 67 (1 2018), pp. 358–380. DOI: https://doi.org/10.1002/hep.29086. URL: https://pubmed.ncbi.nlm.nih.gov/28130846/ (cit. on p. 4).

[6] A. Desai, S. Sandhu, J. P. Lai, and D. S. Sandhu. «Hepatocellular carcinoma in non-cirrhotic liver: a comprehensive review». In: *World journal of hepatology* 11 (1 2019), pp. 1–18. DOI: https://doi.org/10.4254/wjh.v11.i1.1. URL: https://pubmed.ncbi.nlm.nih.gov/30705715/ (cit. on p. 4).

[7] International Consensus Group for Hepatocellular NeoplasiaThe International Consensus Group for Hepatocellular Neoplasia. «Pathologic diagnosis of early hepatocellular carcinoma: a report of the international consensus group for hepatocellular neoplasia». In: *Hepatology (Baltimore, Md.)* 49 (2 2009), pp. 658–664. DOI: https://doi.org/10.1002/hep.22709. URL: https://pubmed.ncbi.nlm.nih.gov/19177576/ (cit. on p. 4).

[8] Torbenson M. S. «Hepatocellular carcinoma: making sense of morphological heterogeneity, growth patterns, and subtypes». In: *Human pathology* 112 (2021), pp. 86–101. DOI: https://doi.org/10.1016/j.humpath.2020.12.009. URL: https://pubmed.ncbi.nlm.nih.gov/33387587/ (cit. on p. 5).

[9] V. Mazzaferro et al. «Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis». In: *The New England journal of medicine* 334 (11 1996), pp. 693–699. DOI: https://doi.org/10.1056/NEJM199603143341104. URL: https://pubmed.ncbi.nlm.nih.gov/8594428/ (cit. on pp. 6, 16).

[10] O. Ciccarelli et al. «Liver transplantation for hepatocellular cancer: UCL experience in 137 adult cirrhotic patients. Alpha-foetoprotein level and locoregional treatment as refined selection criteria». In: *Transplant International* 25.8 (2012), pp. 867–875. DOI: https://doi.org/10.1111/j.1432-2277.2012.01512.x. URL: https://doi.org/10.1111/j.1432-2277.2012.01512.x (cit. on p. 6).

[11] S. Hwang, D.B. Moon, and S.G. Lee. «Liver transplantation and conventional surgery for advanced hepatocellular carcinoma». In: *Transplant International* 23.7 (2010), pp. 723–727. DOI: https://doi.org/10.1111/j.1432-2277.2010.01103.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1432-2277.2010.01103.x (cit. on p. 6).

[12] Q. Lai et al. «Recurrence of hepatocellular cancer after liver transplantation: The role of primary resection and salvage transplantation in East and West». In: *Journal of hepatology* 57.5 (2012), pp. 974–979. DOI: https://doi.org/10.1016/j.jhep.2012.06.033. URL: https://www.sciencedirect.com/science/article/pii/S0168827812005260 (cit. on p. 6).

[13] D. Fuks, S. Dokmak, V. Paradis, M. Diouf, F. Durand, and J. Belghiti. «Benefit of initial resection of hepatocellular carcinoma followed by transplantation in case of recurrence: An intention-to-treat analysis». In: *Hepatology* 55.1 (2012), pp. 132–140. DOI: `https://doi.org/10.1002/hep.24680`. URL: `https://aasldpubs.onlinelibrary.wiley.com/doi/abs/10.1002/hep.24680` (cit. on p. 7).

[14] V. Mazzaferro et al. «Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis». In: *The Lancet. Oncology* 10 (1 2009), pp. 35–43. DOI: `https://doi.org/10.1016/S1470-2045(08)70284-5`. URL: `https://pubmed.ncbi.nlm.nih.gov/19058754/` (cit. on p. 8).

[15] J. Y. Lei, W. T. Wang, and L. N. Yan. «"Metroticket" predictor for assessing liver transplantation to treat hepatocellular carcinoma: a single-center analysis in mainland China». In: *World journal of gastroenterology* 19 (4 2013), pp. 8093–8098. DOI: `https://doi.org/10.3748/wjg.v19.i44.8093`. URL: `https://pubmed.ncbi.nlm.nih.gov/24307805/` (cit. on p. 8).

[16] A. Raj, J. McCall, and E. Gane. «"Metroticket" predictor in a cohort of patients transplanted for predominantly HBV-related hepatocellular carcinoma». In: *Journal of hepatology* 55 (5 2011), pp. 1063–1068. DOI: `https://doi.org/10.1016/j.jhep.2011.01.052`. URL: `https://pubmed.ncbi.nlm.nih.gov/21354447/` (cit. on p. 8).

[17] A. G. Singal, A. Mukherjee, B. J. Elmunzer, P. D. Higgins, A. S. Lok, J. Zhu, J. A. Marrero, and A. K. Waljee. «Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma». In: *The American journal of gastroenterology* 108 (11 2020), pp. 1723–1730. DOI: https://doi.org/10.1038/ajg.2013.332. URL: https://pubmed.ncbi.nlm.nih.gov/24169273/ (cit. on p. 9).

[18] J. D. Liang, X. O. Ping, Y. J. Tseng, G. T. Huang, F. Lai, and P. M. Yang. «Recurrence predictive models for patients with hepatocellular carcinoma after radiofrequency ablation using support vector machines with feature selection methods». In: *Computer methods and programs in biomedicine* 117 (3 2014), pp. 425–434. DOI: https://doi.org/10.1016/j.cmpb.2014.09.001. URL: https://pubmed.ncbi.nlm.nih.gov/25278224/ (cit. on p. 9).

[19] R. Divya and P. Radha. «An Optimized HCC Recurrence Prediction Using APO Algorithm Multiple Time Series Clinical Liver Cancer Dataset». In: *Journal of Medical Systems* 43 (193 2019). DOI: https://doi.org/10.1007/s10916-019-1265-x. URL: https://link.springer.com/article/10.1007/s10916-019-1265-x (cit. on p. 9).

[20] I. Hamamoto, S. Okada, T. Hashimoto, H. Wakabayashi, T. Maeba, and H. Maeta. «Prediction of the early prognosis of the hepatectomized patient with hepatocellular carcinoma with a neural network». In: *Computers in biology and medicine* 25 (1 1995),

pp. 49–59. DOI: https://doi.org/10.1016/0010-4825(95)98885-h. URL: https://pubmed.ncbi.nlm.nih.gov/7600761/ (cit. on p. 10).

[21] A. Cucchetti et al. «Preoperative prediction of hepatocellular carcinoma tumour grade and micro-vascular invasion by means of artificial neural network: a pilot study». In: *Journal of hepatology* 52 (6 2010), pp. 880–888. DOI: https://doi.org/10.1016/j.jhep.2009.12.037. URL: https://pubmed.ncbi.nlm.nih.gov/20409605/ (cit. on p. 10).

[22] W. H. Ho, K. T. Lee, H. Y. Chen, T. W. Ho, and H. C. Chiu. «Disease-free survival after hepatic resection in hepatocellular carcinoma patients: a prediction approach using artificial neural network». In: *PloS one* 7 (1 2012), e29179. DOI: https://doi.org/10.1371/journal.pone.0029179. URL: https://pubmed.ncbi.nlm.nih.gov/22235270/ (cit. on p. 10).

[23] H. C. Chiu, T. W. Ho, K. T. Lee, H. Y. Chen, and W. H. Ho. «Mortality predicted accuracy for hepatocellular carcinoma patients with hepatic resection using artificial neural network». In: *TheScientificWorldJournal* 201976 (2012). DOI: https://doi.org/10.1155/2013/201976. URL: https://pubmed.ncbi.nlm.nih.gov/23737707/ (cit. on p. 10).

[24] G. Qiao, J. Li, A. Huang, Z. Yan, W. Y. Lau, and F. Shen. «Artificial neural networking model for the prediction of post-hepatectomy survival of patients with early hepatocellular carcinoma». In: *Journal of gastroenterology and hepatology* 29 (12 2014), pp. 2014–2020.

DOI: https://doi.org/10.1111/jgh.12672. URL: https://pubmed.ncbi.nlm.nih.gov/24989634/ (cit. on p. 10).

[25] J. Y. Nam et al. «Novel Model to Predict HCC Recurrence after Liver Transplantation Obtained Using Deep Learning: A Multicenter Study». In: *Cancers* 12 (10 2020), p. 2791. DOI: https://doi.org/10.3390/cancers12102791. URL: https://pubmed.ncbi.nlm.nih.gov/33003306/ (cit. on p. 10).

[26] Q. Lai, D. Nicolini, M. Inostroza Nunez, S. Iesari, P. Goffette, A. Agostini, A. Giovagnoni, M. Vivarelli, and J. Lerut. «A Novel Prognostic Index in Patients With Hepatocellular Cancer Waiting for Liver Transplantation: Time-Radiological-response-Alpha-fetoprotein-INflammation (TRAIN) Score». In: *Annals of surgery* 264 (5 2016), pp. 787–796. DOI: https://doi.org/10.1097/SLA.0000000000001881. URL: https://pubmed.ncbi.nlm.nih.gov/27429025/ (cit. on p. 14).

[27] N. Mehta, J. Heimbach, D. Lee, J. L. Dodge, D. Harnois, J. Burns, W. Sanchez, J. P. Roberts, and F. Y. Yao. «Wait Time of Less Than 6 and Greater Than 18 Months Predicts Hepatocellular Carcinoma Recurrence After Liver Transplantation: Proposing a Wait Time "Sweet Spot"». In: *Transplantation* 101 (9 2017), pp. 2071–2078. DOI: https://doi.org/10.1097/TP.0000000000001752. URL: https://pubmed.ncbi.nlm.nih.gov/28353492/ (cit. on p. 15).

[28] A.I. Graham, A. Adair, E.M. Harrison, J. Gordon-Smith, H. Ireland, D. Patel, and S.J. Wigmore. «Implications of diameter and volume-based measurement in assessment criteria for liver transplantation

for hepatocellular carcinoma». In: *Hepatoma Res* 7 (1 2021). DOI: http://dx.doi.org/10.20517/2394-5079.2020.87. URL: https://hrjournal.net/article/view/3843 (cit. on p. 16).

[29] N. B. Ha et al. «Risk factors for hepatocellular carcinoma in patients with chronic liver disease: a case-control study». In: *Cancer causes & control : CCC* 23 (3 2012), pp. 455–462. DOI: https://doi.org/10.1007/s10552-012-9895-z. URL: https://pubmed.ncbi.nlm.nih.gov/22258434/ (cit. on p. 18).

[30] J. Zhang et al. «The threshold of alpha-fetoprotein (AFP) for the diagnosis of hepatocellular carcinoma: A systematic review and meta-analysis». In: *PloS one* 15 (2 2020). DOI: https://doi.org/10.1371/journal.pone.0228857. URL: https://pubmed.ncbi.nlm.nih.gov/32053643/ (cit. on p. 18).

[31] C. Koch et al. «AFP ratio predicts HCC recurrence after liver transplantation». In: *PloS one* 15 (7 2020), e0235576. DOI: https://doi.org/10.1371/journal.pone.0235576. URL: https://pubmed.ncbi.nlm.nih.gov/32614912/ (cit. on p. 18).

[32] P. S. Kamath, R. H. Wiesner, M. Malinchoc, W. Kremers, T. M. Therneau, C. L. Kosberg, G. D'Amico, E. R. Dickson, and W. R. Kim. «A model to predict survival in patients with end-stage liver disease». In: *Hepatology (Baltimore, Md.)* 32 (2 2001), pp. 464–470. DOI: https://doi.org/10.1053/jhep.2001.22172. URL: https://pubmed.ncbi.nlm.nih.gov/11172350/ (cit. on p. 19).

[33] G. P. Guerrini et al. «Value of HCC-MELD Score in Patients With Hepatocellular Carcinoma Undergoing Liver Transplantation». In: *Progress in transplantation (Aliso Viejo, Calif.)* 28 (1 2018), pp. 63–69. DOI: `https://doi.org/10.1177/1526924817746686`. URL: `https://pubmed.ncbi.nlm.nih.gov/29251164/` (cit. on p. 20).

[34] R. Lencioni and L. Crocetti. «Local-regional treatment of hepatocellular carcinoma». In: *Radiology* 262 (1 2012), pp. 43–58. DOI: `https://doi.org/10.1148/radiol.11110144`. URL: `https://pubmed.ncbi.nlm.nih.gov/22190656/` (cit. on p. 20).

[35] R. Lencioni and J. M. Llovet. «Modified RECIST (mRECIST) assessment for hepatocellular carcinoma». In: *Seminars in liver disease* 30 (1 2010), pp. 52–60. DOI: `https://doi.org/10.1055/s-0030-1247132`. URL: `https://pubmed.ncbi.nlm.nih.gov/20175033/` (cit. on p. 20).

[36] Graziadei I. W. «Living donor liver transplantation». In: *Tropical gastroenterology : official journal of the Digestive Diseases Foundation* 28 (2 2007), pp. 45–50. URL: `https://pubmed.ncbi.nlm.nih.gov/18050838/` (cit. on p. 22).

[37] H. Uchiyama et al. «Living donor liver transplantation for hepatocellular carcinoma: results of prospective patient selection by Kyushu University Criteria in 7 years». In: *HPB : the official journal of the International Hepato Pancreato Biliary Association* 19 (12 2017), pp. 1082–1090. DOI: `https://doi.org/10.1016/j.hpb.2017.08.004`. URL: `https://pubmed.ncbi.nlm.nih.gov/28888776/` (cit. on p. 22).

[38] H. R. Liang, C. E. Hsieh, K. H. Lin, C. J. Ko, Y. J. Hung, Y. L. Hsu, and Y. L. Chen. «Living donor liver transplantation for hepatocellular carcinoma beyond the Milan criteria: outcome of expanded criteria in tumor size». In: *BMC surgery* 21 (1 2021), p. 401. DOI: https://doi.org/10.1186/s12893-021-01403-z. URL: https://pubmed.ncbi.nlm.nih.gov/34798847/ (cit. on p. 22).

[39] Lee H. S. «Liver transplantation for hepatocellular carcinoma beyond the Milan criteria: the controversies continue». In: *Digestive diseases (Basel, Switzerland)* 25 (4 2007), pp. 296–298. DOI: https://doi.org/10.1159/000106907. URL: https://pubmed.ncbi.nlm.nih.gov/17960062/ (cit. on p. 22).

[40] A. A. Mokdad, H. Zhu, J. A. Marrero, J. C. Mansour, A. G. Singal, and A. C. Yopp. «Hospital Volume and Survival After Hepatocellular Carcinoma Diagnosis». In: *The American journal of gastroenterology* 111 (7 2016), pp. 967–975. DOI: https://doi.org/10.1038/ajg.2016.181. URL: https://pubmed.ncbi.nlm.nih.gov/27166130/ (cit. on p. 23).

[41] I. S. Yu, S. L. Liu, V.O. Zaborska, T. Raycraft, H.J. Lim, S. Gill, and J. M. Davies. «The impact of geography and center volume on access to care and outcomes in advanced hepatocellular carcinoma (HCC): A retrospective population based study». In: *Journal of Clinical Oncology* 37.15 (2019), e15597–e15597. DOI: 10.1200/JCO.2019.37.15_suppl.e15597. URL: https://doi.org/10.1200/JCO.2019.37.15_suppl.e15597 (cit. on p. 23).

[42] L. Antolini, P. Boracchi, and E. Biganzoli. «A time-dependent discrimination index for survival data». In: *Statistics in medicine* 24 (1 2005), pp. 3927–3944. DOI: https://doi.org/10.1002/sim.2427. URL: https://pubmed.ncbi.nlm.nih.gov/16320281/ (cit. on p. 28).

[43] E.L. Kaplan and P. Meier. «Nonparametric Estimation from Incomplete Observations». In: *Thirty-Second AAAI Conference on Artificial Intelligence* 53.282 (1958), pp. 457–481. DOI: https://doi.org/10.2307/2281868. URL: https://www.jstor.org/stable/2281868 (cit. on p. 29).

[44] U. Katzman J.L. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger. «DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network». In: *BMC Medical Research Methodology* 18 (1 2018), pp. 1471–2288. DOI: https://doi.org/10.1186/s12874-018-0482-1. URL: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0482-1 (cit. on p. 30).

[45] D. Faraggi and R. Simon. «A neural network model for survival data». In: *Statistics in medicine* 14 (1 1995), pp. 73–82. DOI: https://doi.org/10.1002/sim.4780140108. URL: https://pubmed.ncbi.nlm.nih.gov/7701159/ (cit. on p. 31).

[46] H. Kvamme and Ø. Borgan. «Continuous and discrete-time survival prediction with neural networks». In: *Lifetime data analysis* 27 (4 2021), pp. 710–736. DOI: https://doi.org/10.1007/s10985-021-

09532-6. URL: https://pubmed.ncbi.nlm.nih.gov/34618267/ (cit. on pp. 35, 36).

[47] C.N. Yu, R. Greiner, and H.C. Lin. «Learning Patient-Specific Cancer Survival Distributions as a Sequence of Dependent Regressors». In: *Advances in Neural Information Processing Systems* 24 (1 2011), pp. 1845–1853. URL: https://papers.nips.cc/paper/2011/file/1019c8091693ef5c5f55970346633f92-Paper.pdf (cit. on p. 38).

[48] S. Fotso. «Deep Neural Networks for Survival Analysis Based on a Multi-Task Framework». In: *arXiv preprint* (2018), p. 1801.05512. DOI: https://doi.org/10.48550/arXiv.1801.05512. URL: https://arxiv.org/abs/1801.05512 (cit. on p. 39).

[49] V. Mazzaferro et al. «Metroticket 2.0 Model for Analysis of Competing Risks of Death After Liver Transplantation for Hepatocellular Carcinoma». In: *Gastroenterology* 154 (1 2018), pp. 128–139. DOI: https://doi.org/10.1053/j.gastro.2017.09.025. URL: https://pubmed.ncbi.nlm.nih.gov/28989060/ (cit. on p. 42).

[50] C. Lee, W. Zame, J. Yoon, and M. van der Schaar. «DeepHit: A Deep Learning Approach to Survival Analysis with Competing Risks». In: *Thirty-Second AAAI Conference on Artificial Intelligence* 32 (1 2018). DOI: https://doi.org/10.1609/aaai.v32i1.11842. URL: https://ojs.aaai.org/index.php/AAAI/article/view/11842 (cit. on pp. 42, 45).

[51] J.P. Fine and R.J. Gray. «A Proportional Hazards Model for the Subdistribution of a Competing Risk». In: *Thirty-Second AAAI Conference on Artificial Intelligence* 94.446 (1999), pp. 496–509. DOI: https://doi.org/10.2307/2670170. URL: https://www.jstor.org/stable/2670170 (cit. on p. 43).