LU TP 17-nn
June 2017

Analysis of HLA-A, -B and -DR Alleles as Risk Factors for One-Year Mortality in Heart Transplants Using Artificial Neural Networks

Frieder Henning

Department of Astronomy and Theoretical Physics, Lund University

FYTM03

Master thesis supervised by Mattias Ohlsson

LUND
UNIVERSITY

Abstract

The beneficial effects of HLA matching at the HLA-A, -B and -DR loci on recipient short- and long-term survival in heart transplants have been established in a number of studies. The objective of the present study was to evaluate the importance of individual HLA alleles at the HLA-A, -B and -DR loci as risk factors for recipient death within one year after transplant by using artificial neural networks.

Records of 24,838 heart transplants from the UNOS databased consisting of 19 risk variables which include the two recipient and donor HLA alleles at the study loci were analyzed by an ensemble of five multilayer perceptrons in a binary classification task. The importance of the classification variables as predictors was measured in a sensitivity analysis and odds ratios were calculated to identify risk factors for early death. The risk factor calculations for two alleles were validated in a Kaplan-Meier survival analysis. The influence of HLA matching on recipient one-year survival was analyzed in a sensitivity analysis and by risk factor calculations.

The donors HLA-A 23 and HLA-B 53 alleles were identified as risk factors for early death with odds ratios that were comparable to an increased ischemic time of one hour. The HLA-B 27 allele could be associated with decreased risk of early death. These alleles are important predictors to the network ensemble. In the survival analysis, the difference in the survival probabilities of two groups containing 100 transplants with donor HLA-A 23 and HLA-B 27 alleles each was significant ($P = 0.0078$, Log-rank test).

At all loci, one mismatch and at the HLA-A and HLA-DR loci two mismatches were associated with a decreased risk of early death relative to no mismatches. At the HLA-B locus, two mismatches were associated with an increased risk of early death

The network ensemble was able to identify alleles as risk factors and the result was validated. The analysis on the influence of HLA matching gave results that were in conflict with previous studies.

Populärvetenskaplig sammanfattning

Hjärttransplantationer syftar till att rädda liv men kan även vara riskabla ingrepp. Det är mycket som kan gå fel. Det transplanterade hjärtat lider stor risk att avstötas i mottagarens kropp, och en mer detaljerad kunskap om faktorerna som påverkar avstötning har således möjligheten att rädda liv.

Avstötning orsakas av att mottagarens immunförsvar känner av HLA-molekylerna, där HLA står för Human Leukocyte Antigen, på cellytorna hos det transplanterade hjärtat och tolkar dem som kroppsfrämmande. HLA-molekylerna delas in i två klasser och i varje klass ingår tre klassiska HLA molekyler. De klassiska HLA-molekylerna skiljer sig i sin form genom fysiska variationer, vilka har sitt ursprung i människans genuppsättning där HLA-molekylerna är kodade i HLA-komplexet. I HLA-komplexet motsvarar de olika HLA-molekylerna positioner, så kallade loci, och för varje locus finns ett antal genvarianter. Genvarianterna kallas alleler och en individ kan ha upp till tolv olika HLA-alleler.

Inom teorin för organavstötning spelar HLA-molekylerna en central roll. De positiva effekterna på mottagarens överlevnadstid efter en transplantation när mottagarens och donatorns HLA alleler matchar är allmänt vedertagna.

Det som i denna studie undersöks är om risken för avstötning inom ett år efter hjärttransplantationen ökar när vissa alleler förekommer hos antingen mottagaren eller donatorn. Detta innebär att data om hjärttransplantationer undersöks med en högre upplösning jämfört med tidigare studier.

I projektet utvärderas data från fler än tjugotusen hjärttransplantationer med hjälp av artificiella neurala nätverk som är en typ av artificiell intelligens. Artificiella neurala nätverk kan lära sig att självständigt känna igen mönster i data med många variabler och har därför blivit ett allt populärare verktyg inom dataanalys de senaste åren.

Contents

1 Introduction

A heart transplant is the surgical procedure in which the diseased heart of a patient (recipient) is replaced by a healthy heart of a deceased donor in order to increase the recipients expected lifespan. Heart transplant surgery is a well established treatment for patients with weakened or damaged hearts and patients where other treatments have failed [1]. Since the first heart transplant in the 1960s, the patients' median survival time after transplant has been increased and keeps increasing due to progress in several different areas which include the refined selection of recipients and donors and improved medical therapies that accompany a heart transplant [2].

In spite of the tremendous advances, a heart transplant still remains a hazardous intervention. In part, this is due to the risk of the donor heart being rejected or injured by the body of the recipient, i.e. graft rejection and graft failure. Graft rejection and sometimes even graft failure is caused by the immune response of the patient against the transplanted heart (graft). Often the immune response leaves the graft dysfunctional with potentially lethal consequences for the recipient [3].

The immune response that is launched against the graft is triggered by the Human Leukocyte Antigen (HLA) molecules. HLA molecules are cell surface molecules that can be found on virtually any cell in the human body and are thus even present on the transplanted graft. The presence of foreign HLA molecules in the body of the recipient is recognized by cells of the immune system and an immune response against the graft is released [4].

Two different classes of HLA molecules are distinguished and both HLA classes contain three classical HLA molecules each. For any of the classical HLA molecules there exist different versions that differ in shape. On the genetic level, the diversity of HLA molecules that can be observed in the human population corresponds to a large number of genetic variations in the gene complex that encodes the HLA molecules (HLA complex). On the HLA complex, every HLA molecule is encoded at a certain position (locus) and for every gene locus there exist different numbers of genetic variations (alleles) that determine the shape of the HLA molecule. Every human inherits one allele for each locus from each of its parents and both alleles are expressed equally [4].

In previous studies, it was found that matching the HLA alleles of the recipient and donor (HLA matching) at the HLA-A, -B and -DR loci in a heart transplant improves recipient median survival after transplant. In particular, matching of the HLA-DR alleles improves one-year recipient survival significantly [5]. In contrast to

HLA matching, to our knowledge, no studies on the influence of individual alleles on recipient survival have been done.

In the present study, the influence of the presence of the alleles of the HLA-A, -B and -DR loci in heart transplants on recipient one-year survival was studied in a risk factor analysis using artificial neural networks (ANNs). In the analysis, records of more than twenty thousand heart transplants were divided into two classes depending on whether the recipient was still alive one year after transplantation. The transplant records were classified by an ensemble of five ANNs based on the recipient and donor HLA alleles at the study loci and 13 additional classification variables that were expected to be linked to the influence of HLA molecules on the outcome of heart transplants. In the risk analysis, original variable values were altered and the resulting changes in the probability for one-year mortality for different variable values were compared. In similar calculations, the influence of HLA matching at the same loci on recipient one-year survival was evaluated.

In the first chapter of this report, some basic theory of ANNs and the medical background on the importance of HLA molecules in heart transplants is given. The second chapter is concerned with the methods that are of particular importance in data preprocessing and model selection in this project. Further, methods for evaluating the importance of variables as predictors, risk factors and differences in survival probabilities in two groups are introduced. In the third chapter, the results of the application of these methods to data that contains full information about the HLA alleles present in heart transplants is presented and the results of the sensitivity analysis and risk factor evaluation for HLA matching data are shown. In the following section, the quality of the results is reviewed. This discussion includes comments on the network ensemble members and the signal of HLA variables in the dataset. Related to this signal, replacing missing data values and the usage of external data is discussed. Further possible practical implications of the result are mentioned and open questions that follow from the present study are touched. The section is completed by briefly discussing the result of the HLA matching analysis. In the last section, conclusions that can be drawn from the previous work are summarized.

2 Theory

In this section, basic theory on the functioning of ANNs and the role of HLA molecules in graft rejection is given. The ANN theory is mainly focused on a particular architecture, namely multilayer perceptrons for binary classification and both technical aspects and general terms that are important in the following chapters are presented. To give a basic understanding of the importance of HLA molecules in heart transplants, a short overview of heart transplants is given, the processes that lead to graft rejection where HLA molecules are involved are described and the origin of the diversity in HLA molecules is explained by briefly presenting the HLA complex.

2.1 Artificial Neural Networks

In a classification problem it is desired to mark observations by a label from a set of M different categories using a classifier. ANNs can be used as probabilistic classifiers that map an input vector onto M numbers between zero and one that are interpreted as the probabilities of a given input to belong to the M categories. Classification problems with $M = 2$ are called binary classification problems. In binary classification problems the network output reduces to a single number that gives the probability of the input for belonging to one of the two classes. From this number, the probability of the input for belonging to the other class can be deduced. The classification is enabled by the ability of ANNs to learn and different learning strategies exist that allow for turning an ANN into a meaningful classifier. In supervised learning, training data consisting of training inputs with known target values is used to train an ANN that returns the correct labels for as many training inputs as possible. The goal of supervised learning is to create ANNs that make relevant label predictions on new inputs that have not been part of the training (i.e. generalization) [6].

ANNs are networks of artificial neurons that are connected to each other through weights. Depending on the type of problem that is to be solved, varying ANN architectures are employed. An architecture that is often used in classification tasks is the multilayer perceptron (MLP) [7].

The basic unit of an ANN is the artificial neuron. A conceptual representation of an artificial neuron is shown in figure 2.1. From the left hand side an input vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is passed to the neuron and weighted with the weights $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$. The weighted sum $h(\mathbf{x}, \boldsymbol{\omega}) = \mathbf{x} \cdot \boldsymbol{\omega}$ is calculated and the bias b is added. The result is passed to the activation function $\varphi(h, b)$ to generate the output y [7].

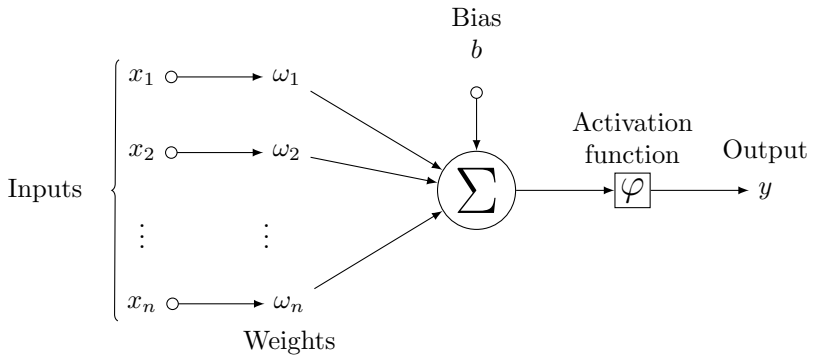


Figure 2.1: The artificial neuron. From left to right, the vector components $x_1, x_2, \ldots, x_n$ of an input vector $\mathbf{x}$ are weighted by the weights $\omega_1, \omega_2, \ldots, \omega_n$, summed and the bias b is added. The result is passed to the activation function φ to generate the output y.

Artificial neurons are linear classifiers. The weights and bias define a hyperplane, the so called decision boundary, that divides input space into two regions and allows classification of an input vector $\mathbf{x}$ according to the region its output is associated with. A problem is said to be linearly separable if all points in the problem can be separated correctly by a hyperplane [7].

MLPs are interconnected sets of artificial neurons. The structure of a MLP is shown in figure 2.2. In a MLP, three layer types are distinguished. These are the input layer, the hidden layer and the output layer. The nodes in the input layer are determined by the input vector $\mathbf{x}$ whereas the nodes of the hidden layers and the output layer are made up of artificial neurons. From the input layer, the input signal is fed forward to the nodes in the first hidden layer. For each node in the hidden layer, the signal from the previous layer is weighted by an individual weight vector before the outputs of the nodes are calculated and propagated forward to the next layer. In this way, the signal is passed forward through the MLP until the output layer is reached where the classification result is calculated. The number of hidden layers and the number of nodes per layer in a MLP is problem dependent. MLPs for binary classification problems typically possess a single output node [7].

The nodes in a MLP act as feature detectors and decision boundaries of MLPs can be more complex compared to those of single artificial neurons. The shape that the decision boundary can assume is dependent on the number of hidden layers [7].

For the functioning of MLPs, the choice of the activation functions of the nodes in the hidden and output layer is crucial for the type of classification tasks that can be solved and for the interpretation of the network output. The standard activation function of the output node of a MLP for binary classification is the logistic activation function

$$\varphi(x) = \frac{1}{1 + e^{-x}} \ . \tag{2.1}$$

The output y of such a MLP is a real number between 0 and 1. For a given input, y is interpreted as the probability of the input to belong to one of the two classes [7].

MLPs whose nodes have nonlinear activation functions in the hidden layers are able to solve linearly non-separable classification problems. A typical choice is a

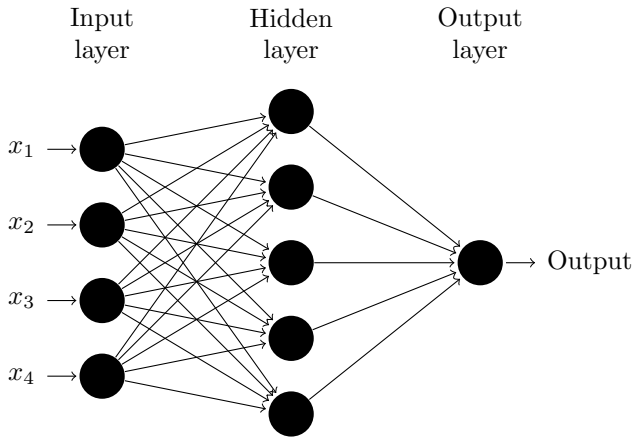


Figure 2.2: Conceptual representation of a MLP with four input nodes, one hidden layer and one output node. Generally, the number of input nodes, hidden layers and nodes in the hidden and output layers is problem dependent. The output of the input layer corresponds to the input vector. The vector components are weighted by individual weight vectors and passed to the neurons in the first hidden layer. The output of the nodes in the hidden layer are passed to the output layer where the network output is calculated.

sigmoid function, e.g. equation (2.1), but the rectifier, defined as

$$\varphi(x) = \max(0, x) \ , \tag{2.2}$$

is a possible activation function as well [8].

In the context of MLPs, learning means the ability of the network to autonomously find the set of weights $\boldsymbol{\omega}$ that returns the correct target value for a given training input. Learning takes place in the training stage where the MLP uses training data consisting of inputs $\mathbf{X} = (\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N))$ and target values $\mathbf{d} = (d(1), d(2), \ldots, d(N))$ to learn the correct labeling. In the training stage, the training data is repeatedly passed to the ANN. For the training inputs, small changes of the weights are calculated in order to find the optimal weight configuration that returns the correct target value for as many training inputs as possible. The changes in the weights are calculated to minimize an error function E that is introduced as measure for the difference between the network output and target values. The changes of weights that minimize E are calculated by means of enhanced gradient descent methods and the gradient of E is calculated using the backpropagation algorithm [7].

Before training, the MLP is initialized with small, random weights and the training data is split into batches, i.e. data segments of a certain size. The inputs of a given batch are passed to the ANN, the outputs are calculated and compared to the target values using E. For every batch member a weight update that minimizes E is calculated for all weights. After the weight updates for all batch members have been calculated, the weights in the ANN are updated by the averaged batch member updates. A training epoch is completed when all batches have been passed to the ANN once. In the next training epoch, new batches are generated [7].

The error function that is commonly used in binary classification problems is the

binary cross-entropy error function. The binary cross-entropy function is given by

$$E = -\sum_{n=1}^{N} \left[d(n) \log y(n) + (1 - d(n)) \log \left(1 - y(n)\right) \right] \tag{2.3}$$

and can be derived from probability theory [7]. It is non-negative as the individual terms in the sum are negative and the sum is multiplied by -1. Assuming the labels $d(n)$ can assume the values 0 and 1, for $y(n) \approx d(n)$ for all n, the the first term of the sum equals approximately zero as it is multiplied by zero when $d(n) = 0$ while the logarithm approximately vanishes for $y(n) \approx 1$. The first factor in the second term in the sum vanishes for $d(n) = 1$ while the second factor vanishes for $y(n) \approx 0$. It can be concluded that the error function is minimized when the outputs are as close as possible to their target values [7].

Ideally, after training, the MLP is able to make correct statements on some validation data that has not been part of the training. The performance on the validation data is an estimate of the MLPs generalization performance. To optimize the generalization performance, ANN often need to be prevented from overfitting in the training stage. Overfitting occurs when the ANN learns to reproduce the noise in the data rather than learning recurring patterns in the input. To ensure good generalization performance, overfitting needs to be controlled. For this purpose network regularization techniques are introduced [7].

Network regularization techniques are schemes that aim to prevent ANNs from overfitting and thus to improve the generalization performance. Common regularization techniques that use different approaches to network regularization are L2 regularization, dropout and weight normalization.

In L2 regularization a regularization term is added to E and the total error function becomes

$$\tilde{E} = E + \frac{1}{2}\lambda \sum_{i} \omega_i^2 \ . \tag{2.4}$$

Here, E is the unmodified error function (2.3). The regularization term is the product of the sum of the squared weights and a regularization parameter λ that specifies the contribution of the regularization term to the overall error. In ANNs that use L2 regularization large weights are penalized and weight vectors of L2 regularized ANNs tend to be more diffuse. This, ideally, makes ANNs less dependent on single input dimensions and forces it to emphasize recurrent patterns in the training data instead [7].

Another common regularization approach is dropout. In the training of networks with dropout, a certain fraction p of randomly chosen nodes are set to zero for every input. Only the remaining nodes have their weights updated. After training, the weights are scaled down to account for the presence of all neurons in the final ANN. Dropout makes the training of the network more robust to changes in the data as no neuron can rely on the presence of the other [9].

Weight normalization is a regularization method that fixes the absolute value of the weight vectors in ANNs. Weight vectors are forced to assume a certain length g by demanding

$$\tilde{\boldsymbol{\omega}} = \frac{g}{|\boldsymbol{\omega}|}\boldsymbol{\omega} \ . \tag{2.5}$$

The normalized weight vector $\tilde{\boldsymbol{\omega}}$ is the product of the original weight vector $\boldsymbol{\omega}$ and the weight normalization parameter g divided by the Euclidean norm of the original weight vector $|\boldsymbol{\omega}|$. Weight normalization has been reported to improve the performance of gradient descent training [9].

K-fold cross-validation is a validation technique that is used to estimate the generalization performance of an ANN. In K-fold cross-validation, a dataset $\mathbf{X}$ is split into K subsets of approximately equal size. These K data folds are arranged into K different pairs of training data and validation data, where each fold is used as validation data once while the other $K-1$ folds serve as training data. With this arrangement, K ANNs with K validation performances are trained. The generalization performance of an ANN is estimated as the average of the validation performances for the K validation folds [7].

Besides the regularization methods that are described above, there are many other techniques that can be introduced to improve convergence during training or generalization performance. In particular, to optimize generalization performance further, often network ensembles are used instead of single ANN and input data is preprocessed to facilitate its handling by the ANN [7].

Network ensembles are collections of ANNs whose outputs are combined to obtain a more stable or even improved performance of the ensemble compared to the performances of its individual members. In network averaging, the ensemble output is a linear combination of the outputs of the members. The ensemble can be made up of ANN with different architectures or ANN that have the same architecture but are trained on differing data. To achieve robust results, it is desirable to construct network ensembles with members that are as versatile as possible [7].

Bootstrapping is a resampling technique that is used to increase the diversity of ANNs that are trained on the same dataset. A bootstrap sample of a dataset with N data points is a dataset of size N with replacement, i.e. random data points appear multiple times in the bootstrap sample and an equal number of data points is removed. ANNs that are trained on bootstrap samples of the same dataset are thus exposed to slightly differing information in training [7].

Data preprocessing is sometimes required to maximize the amount of information that ANNs can extract from data. For instance, when input variables differ in the ranges that are covered by their values, it is desirable to scale them such that the values cover similar ranges. Continuous variables are often normalized by rescaling them to follow the standard normal distribution with zero mean and standard deviation one. Discrete variables can be translated into strings of zeros and ones [7].

The architecture, activation functions, regularization techniques and their hyperparameters, etc. of an ANN are generally chosen as to maximize its performance. Typically, these properties are found in experiments and their application is justified by the improved functionality of the ANN. Different strategies have been used to estimate the optimal properties and parameters of ANNs for a given problem. In a grid search possible combinations of parameter sets and properties are used to find the optimal ANN. This requires training of many networks where, depending on the parameters and properties that are searched, many networks might be very similar to each other. In an alternative approach the parameters are found in a random search [10].

2.2 Heart Transplants

A heart transplant is a surgical intervention in which the heart of the recipient is replaced by the heart of another human (donor) who has been declared brain dead. It is the definitive treatment for end-stage heart failure patients. Worldwide, 4,000 to 5,000 heart transplants are carried out annually [3]. Nowadays, about 90% of the patients survive the first year after transplant, 70% survive five years after transplant and the median survival is somewhat higher than ten years [2].

The allocation of donor hearts follows certain criteria. These are to ensure a fair distribution of the available hearts among the patients as the number of patients exceeds the number of available grafts. Besides the health condition of the recipient and his or her time on the waiting list, matching criteria are used to achieve a maximum survival chance for the recipient. These matching criteria are ABO blood group compatibility, weight matching, age matching and sex matching. HLA matching is no criterion in graft allocation [2].

A constraint that limits the viability of donor-recipient matching is the ischemic time. The ischemic time is the time between the graft removal from the donor and the transplant. Generally, it should not exceed four to five hours and imposes thus a geographical constraint on possible recipient-donor pairs [2].

Possible complications after a transplant are graft failure and graft rejection. Graft rejection is the destruction of the graft in the recipients body through the adaptive immune response [4]. The adaptive immune response can even result in an injury of the graft leading to graft failure [3].

The adaptive immune response is the reaction on the attack of a pathogen, e.g. a virus or bacteria, that is executed by the cells of the adaptive immune system and aims for the removal of the pathogen. The adaptive immune response is initiated by T and B cells that have been activated in the interaction with an HLA molecule in complex with a peptide of non-self origin (antigen). These interactions are enabled by receptors that T and B cells are provided with [11].

A class of cells that is an important part of the adaptive immune system are the professional antigen presenting cells (APCs), e.g. dendritic cells (DCs) and even B cells. APCs are specialized in absorbing pathogens and degrading them into peptides. The peptides are subsequently presented to the cell exterior in complex with HLA molecules that are sitting on the membrane of the APCs [11].

There are two T cell subpopulations that interact with two classes of HLA molecules. The HLA molecule classes are expressed on different cell types. HLA class I molecules interact with CD8 T cells and are expressed on practically all nucleated cells. HLA class II molecules interact with CD4 T cells and are expressed on APCs. The two HLA molecule classes present peptides of two origins that give rise to different immune responses [12].

HLA class I molecules present peptides of endogeneous origin, i.e. peptides that are derived from the interior of a cell. Inside nucleated cells, proteins are routinely fragmented into peptides and then passed to HLA class I molecules that are sitting on the cell membrane. The HLA class I molecule and the peptide form a complex which is exposed to the cell exterior. The recognition of an antigen/HLA class I complex by CD8 T cells leads typically to a cytotoxic response that induces the death of the

host cell [12].

HLA class II molecules present peptides of exogeneous origin that are derived from extracellular pathogens that have been internalized by an APC. On the recognition of a antigen/HLA class II complex, CD4 T cells stimulate the production of CD8 T cells and activate B cells that in turn start to secrete antibodies that can remove the attacking pathogen [12].

T and B cells can be found in high concentrations in the lymph nodes and the spleen which make up the secondary lymphoid tissue but a fraction of them does even circulate in the lymphatic system. In the secondary lymphoid tissue, pathogens are filtered from blood and lymph. Moreover, mobile APCs move to the secondary lymphoid tissue upon absorption of a pathogen [12].

The process in which donor cells are recognized by recipient T and B cells after a transplant is called allorecognition and there are two ways through which allorecognition can take place. These are the direct and the indirect pathway of allorecognition [13].

In the direct pathway, cells from the graft migrate to the secondary lymphoid tissue where the non-self HLA molecules on the cell surfaces are recognized by T cells. This scenario is often associated with the mobile non-self DCs that reside in the graft. Upon recognition of non-self HLA molecules, the immune response is launched by the T cells [13].

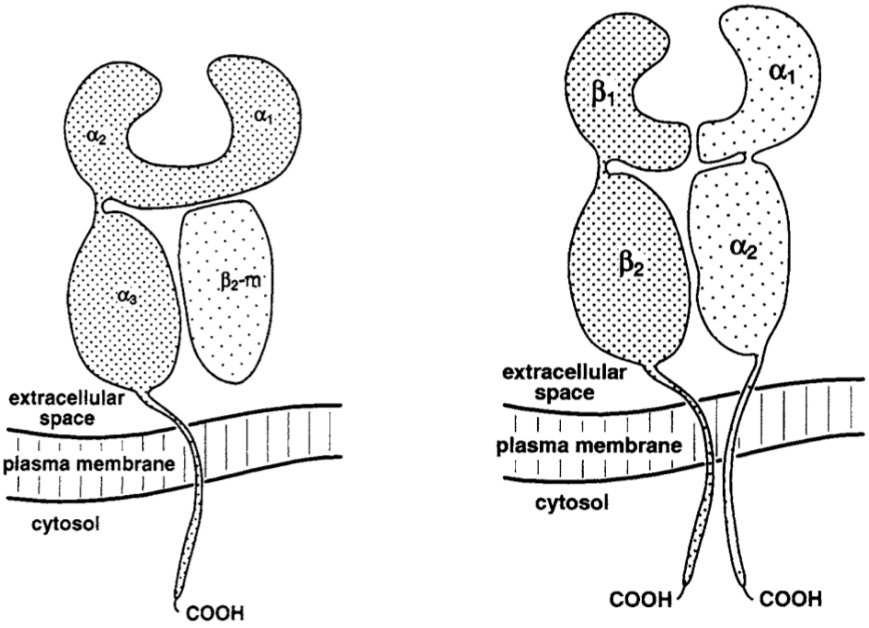
In the indirect pathway, self APCs present preprocessed donor cells to the CD4 T cells. Possible routes for APCs to come into contact with cells of non-self origin go via donor cells that migrate to the secondary lymphoid tissue where they are absorbed by APCs or alternatively via APCs that penetrate into the graft, pick up a cell and migrate further into the secondary lymphoid tissue where the cell fragments are presented to CD4 T cells [13].

In allorecognition, often the presence of foreign HLA molecules is recognized before the presence of other foreign molecules. In part, this is due to the vast number of variations that exist for some HLA molecules which make it highly likely that recipient and donor cells differ in the HLA molecules on their surfaces. The variety in HLA molecules is a consequence of a large number of alleles that exist for each locus at the HLA complex [14].

The HLA molecules that are most important in allorecognition are the HLA class I and HLA class II molecules. Molecules of both classes are subdivided into classical and non-classical molecules. Only the classical molecules are directly involved in the interaction with T and B cells. In spite of the allelic variations, the classical molecules of each class have similar structures [15].

The classical HLA class I loci are the HLA-A, -B and -C loci. The structure of classical class I molecules is shown in figure 2.3a. The extracellular part of the HLA class I molecule consists of a α polypeptide chain and a β chain. The α polypeptide chain is attached to the transmembrane anchor that binds the HLA molecule to the cell membrane. Only the heavy chain and the anchor are encoded on the HLA complex [15].

The heavy chain consists of three domains. These are the α_1, α_2 and α_3 domains. The α_1 and α_2 domains combine to constitute the peptide binding groove. The shape of the binding groove determines the shape of peptides that can be presented by



(a) HLA class I molecule. HLA class I molecules consist of an extracellular part and a transmembrane anchor. The extracellular part is built up of a α polypeptide chain and a β chain. The α polypeptide chain is made up of three domains whereof the α_1 and α_2 domain make up the molecule binding groove where the peptides are clamped and presented to the exterior. The polymorphism of HLA class I molecules is often due to differences in the α_1 and α_2 domains.

(b) HLA class II molecule. HLA class II molecules consist of an extracellular part and two transmembrane anchors. The molecules are built up of one α and one β chain that consist of two domains each. The α_1 and β_1 domains make up the binding grove of HLA class II molecules. The polymorphism of HLA-DR molecules is due to differences in the β chain. The polymorphism of HLA DP and HLA-DQ molecules can even be due to differences in the α chain.

Figure 2.3: HLA class I and HLA class II molecules [15].

the HLA molecules. The polymorphism of HLA class I molecules is often due to differences in the α_1 and α_2 domain. The number of classical HLA class I molecules can be found in table 2.1 [15].

Table 2.1: Number of classical HLA class I molecules.

Locus	No of molecules
HLA-A	28
HLA-B	62
HLA-C	10

The classical HLA class II loci are the HLA-DP, -DQ and -DR loci. The structure of an HLA class II molecule is shown in figure 2.3b. Class II molecules consist of a α and a β chain that are composed of two domains each. Both chains are anchored to the membrane. The α_1 and β_1 domain make up the peptide binding groove [15].

The polymorphism in HLA-DR molecules is due to differences in their β chain whereas the α chain is monomorphic. In HLA-DP and -DQ molecules, the polymor-

phism can even be due to differences in the α chain. The number of HLA class II molecules can be found in table 2.2 [15].

Table 2.2: Number of classical HLA class II molecules.

Locus	No of molecules
HLA-DP	6
HLA-DQ	9
HLA-DR	24

The genome of any individual contains two HLA alleles for each locus that are expressed codominantly. Consequently, every individual can have up to 12 different classical HLA molecules expressed on the cell surfaces [14].

In the period from 2002 to 6/2012, 36% of recipient deaths within 30 days after transplant were due to graft failure and 4% due to graft rejection. This makes graft failure the most important cause of death within the first 30 days after transplant while graft rejection is third most important. In the period from 31 days to one year after transplant, 17% percent of the recipients died by graft failure and 8% by graft rejection. This makes graft failure the second most important cause of death for the time between 31 days and one year after transplant and graft rejection the third most important cause of death for the same interval [16].

To prevent graft rejection and immune injuries, recipients undergo an immunoresponse therapy previous to the heart transplant. The immunoresponse therapy aims to deplete immune cells and to weaken the adaptive immune response. Advances in the development of the immunoresponse therapy have lead to an decreasing number of graft rejections over the last decades [13].

3 Method

In this section, the database and the classification variables that were used in this project are introduced. The procedures for data preprocessing are presented with particular focus on the encoding of the HLA loci. The area under the receiver operating characteristic curve as measure for evaluating the classification performance of binary classifiers is explained and the construction of the network ensemble is described. Further, a method that uses differences in the error function of ANNs for ranking variables by their importance as predictors and how it was applied in this project is presented. Odds ratios are introduced as association measure between an exposure and an outcome that allows to identify risk factors and it is described how odds ratios were calculated in the present study. Finally, it is explained how survival probabilities in two groups can be compared using a Kaplan-Meier survival analysis.

3.1 Database

The database used in this project was collected by the United Network for Organ Sharing and contained records of heart transplants that were carried out in the United States since 1987. The original dataset consisted of 52,515 heart transplants some of them being complete. More information about the background of the database can be found in [17].

The transplant records contained the date of death of the recipient after transplant. This was used to divide the transplants into two classes depending on whether the recipient was still alive one year after transplant. The cause of death was not specified in the data.

In this project, 19 predictor variables were considered as predictors to classify the transplants. Of the 19 variables, 15 were related to recipient-donor matching. The matching variables were ethnicity (Eth), age, gender, weight (Wgt) and the two recipient and donor HLA-A, -B and -DR alleles. One additional matching variable was ABO blood type matching between recipient and donor (ABO Match). Three variables described the recipient only. These were immune activity (High PraMR), whether the recipient had an infection at the time of transplant (infect) and the number of graft transplants the recipient had undergone previous to the heart transplant (Prev TX). The last variable that was kept was the ischemic time (Isch Time). The values of the study variables can be found in appendix A. The variables other than the HLA alleles were selected as they were expected to be linked to the impact of HLA molecules on the outcome of heart transplants.

To study the effect of individual HLA alleles and HLA matching separately, two datasets were created from the UNOS database. In the allele dataset, detailed information about recipient and donor HLA alleles at the study loci was contained. In the matching dataset, the number of mismatches at the study loci for each transplant was considered.

3.2 Data Preprocessing

In the data preprocessing, transplants with missing data were removed from the dataset. Further, transplants where either the recipient or donor was younger than 15 years were excluded from the study as it was assumed that the adaptive immune system is not fully developed in individuals below this age.

Continuous variables were normalized. Discrete variables were encoded by strings of zeros and ones. The length of the string that represented a given variable was equal to the number of values that the variable can assume. Strings that represented non-HLA variables contained a single one and zeros otherwise.

In the allele dataset, the two alleles of the study loci of the donor and recipient were encoded in a single string that contained either a single one or two ones. The scheme that was used for encoding the HLA profile at a given locus is shown in figure 3.1. In the first step, the two alleles were encoded separately into strings that contained a single one only. The strings were then combined into one string. When the HLA profile was made up of two different alleles, the combined string was the sum of the individual strings and contained two ones and zeros otherwise. When the HLA profile was made up of one allele only, the combined string was given by the string that encoded this particular allele. This procedure was used to ensure that the two alleles at every locus were treated equally by the MLPs.

In the matching dataset, matching variables for the three loci were introduced. Each locus was represented by a string of length three that could represent the values two mismatches, one mismatch and no mismatch. The value no mismatch was defined relative to the recipient: When the recipient had alleles X and Y at some locus where the donor had allele X only, this corresponded to no mismatch.

3.3 Performance Measures

The performance of a binary classifier can be measured by its sensitivity and specificity. The sensitivity or true positive rate (TPR) is the probability for a positive data point to be classified as positive. The specificity is the probability of a negative data point to be classified as negative. The false positive rate (FPR) of a classification can be related to the specificity as it is the probability for a negative data point to be incorrectly classified as positive. The FPR is therefore equal to $1 - \text{specificity}$ [18].

The receiver operating characteristic (ROC) curve is the plot of the classifiers TPR against its FPR for varying classification thresholds. An example of a ROC curve is shown in figure 3.2. The ROC space is given by a square with side length one. The point $(0, 1)$ characterizes a perfect classification as it represents a classification with sensitivity and specificity equal one. The ROC curve of a high quality classifier lies therefore close to this point. In contrast, a classifier with a ROC curve that follows the diagonal, i.e. a classifier whose TPR is equal to its FPR for all classification thresholds, is unable to discriminate the classes. ROC curves that lie below the diagonal can be

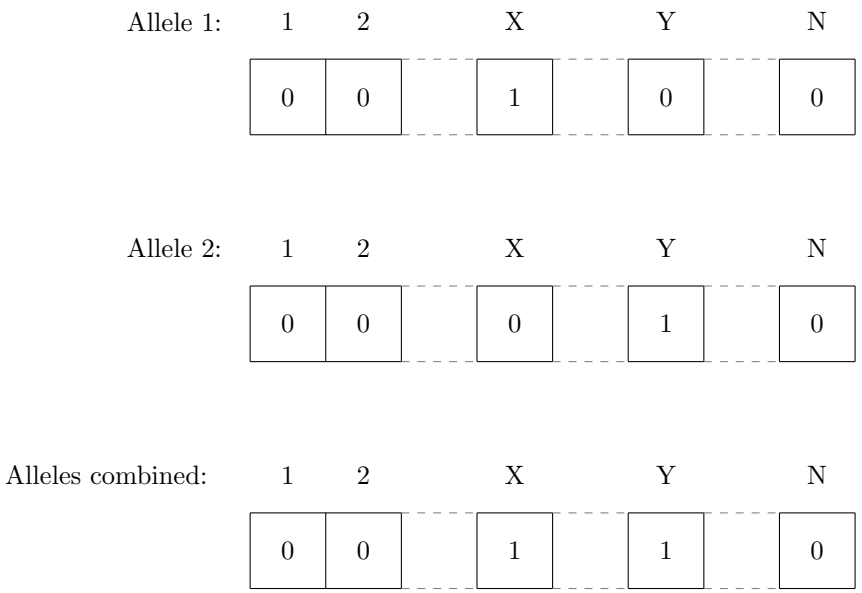


Figure 3.1: The encoded HLA profile of a human with alleles X and Y at a locus with N alleles. When the HLA profile contained two different alleles, the encoded profile was the sum of the strings that encoded the individual alleles. When the profile consisted of one allele only, the encoded profile was given by the string that encoded this particular allele.

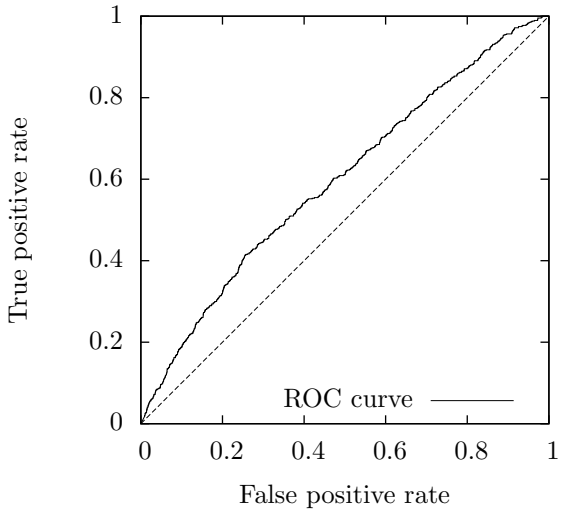


Figure 3.2: The ROC curve of a classification that was performed by a MLP that is later used in the project. The ROC curve is the plot of the TPR versus FPR of a classification at varying classification thresholds. The ROC space is a square with side length one where the point (1,0) corresponds to a perfect classification. In the figure, the diagonal that indicates a random classifier is shown as dashed line. The ROC curve lies above the diagonal. In this example the ROC curve has an AUC of 0.597.

inverted by inverting the classification. The ROC curve is independent of the label distribution in the data and independent of the cost function [18].

The area under the ROC curve (AUC) is often used to evaluate the performance of a binary classifier [18]. The AUC is a number between 0.5 and 1 and can be

interpreted as the probability for a classifier to rank a randomly chosen positive data point higher than a randomly chosen negative data point [19].

3.4 Model Selection

For the classification of the transplants, a network ensemble with five members was used. All members were MLPs. In the construction of the ensemble members, certain properties that were assumed to be optimal were shared by all MLPs whereas some parameter values for the MLPs were determined in a grid search.

Properties that were shared by all ensemble members were the regularization techniques that were used, normalization parameter g, optimizer, weight initialization and batch size. Further the length of the training was chosen such that the performance measure had converged at the end of the training. The choice of these properties was based on fastest convergence of the performance measure for a particular MLP architecture with fixed number of hidden layers and nodes in the hidden layer.

Parameters that were determined in a grid search were the number of hidden layers and the number of nodes in the hidden layers, the dropout with differing p-values in the input and hidden layers and the regularization parameter λ.

In the grid search for the parameter values, the allele dataset was used. The same MLP architectures that perform best on this dataset were assumed to be appropriate for classifying the matching data as well.

In this project, 5 fold cross-validation was used to estimate the generalization performance of the five ensemble members. The MLPs were trained on bootstrap samples of the five training datasets that were obtained by the K-fold cross-validation method. Consequently, in the calculations, every patient was classified by five MLPs. The average of the five predictions was taken to be the classification probability and their standard deviation was taken to be the uncertainty in the classification.

The code was written in Python using the Keras deep learning library [20].

3.5 Sensitivity Analysis

In the context of classification problems, a sensitivity analysis is the ranking of variables by their importance as predictor. Often several different methods of varying complexity can be used to calculate a sensitivity ranking for a given classifier. A possible complication that can arise when performing a sensitivity analysis are correlations between variables as correlations make it hard to quantify the importance of variables individually. In spite of the difficulties, it is interesting to perform a sensitivity analysis in order to gain insight into the data and the working of a classifier [21].

A particularly simple approach to a sensitivity analysis for ANNs is to calculate the difference in the error function for the cases where full information about a given variable is contained in the data and the case where this information is removed [21].

In this project, removal of information meant the replacement of the values of a given variable in the validation data by the average of the same value in the training data. The sensitivity to a variable was then calculated as the difference of the error function of the MLP for validation data with information removed and the error function of the MLP for complete validation data. The MLPs were not re-trained on the data with removed information which bears the potential of corrupting the sensitivity analysis [21]. However, as there were many variables, re-training the ANNs

would have become an extremely time-consuming task.

Sensitivities of variable values for the individual ensemble members were calculated for the five validation folds and normalized. The averaged normalized sensitivities were taken to be th sensitivity of a variable. The standard deviation in the averaged normalized sensitivities was taken to be the uncertainty in the sensitivity.

3.6 Odds Ratios

In medical applications, odds ratios (ORs) are used as association measure between an exposure and an outcome. The calculation of ORs is based on the odds for the event in the case where exposure is given and the case where there is no exposure. For binary variables, ORs can be defined unambiguously. In contrast, for discrete, non-binary variables and for continuous variables, the exact meaning of exposure and no exposure needs to be defined before ORs can be calculated [22].

The odds for an event is a measure for the likelihood of the event to happen. It is defined as the ratio of the probability of the event to happen to the probability of the event not to happen,

$$\text{odds} = \frac{p}{1-p} \ . \tag{3.6}$$

With odds for an event larger than one, the likelihood that a given event occurs is higher than the likelihood of the event not to occur.

In medical applications, the OR is an association measure between a certain exposure and an outcome of interest. When p_1 is the probability for the outcome in a group with the given exposure and p_0 is the probability for the same outcome in the group without exposure, the OR is defined as,

$$\text{OR} = \frac{p_1/(1-p_1)}{p_0/(1-p_0)} \ . \tag{3.7}$$

The OR has a simple interpretation. An OR greater than one means that the outcome is associated with the exposure as the likelihood for the outcome is larger in this group compared to the group without exposure. Exposure is then said to be a risk factor for the outcome. Vice versa, an OR less than one means that the event is associated with no exposure. An OR that is equal to one is inconclusive and the outcome is associated with neither group [22].

The above definition can easily be applied to binary variables where exposure and no exposure can be defined unambiguously. When it is desired to calculate the OR of non-binary variables, an exact definition of exposure and no exposure is required. The most important non-binary variables in this project were the HLA loci that encode the exposure to the alleles. Other non-binary variables were the recipients and donors ethnicity, the number of transplants previous to the heart transplant and the continuous variables.

Exposure to a given allele was defined by the string that contains the given allele only. Exposure to allele X at a HLA locus with N alleles was thus represented by a string of length N that contained $N-1$ zeros and a one at position X. The corresponding string is shown in figure 3.3. The absence of exposure could be represented by any string of length N that contained one or two ones, but none of them at position X.

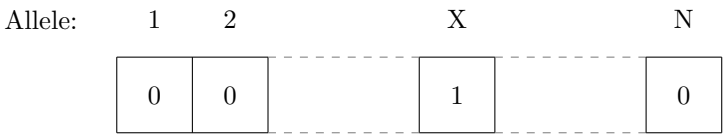


Figure 3.3: The string that represented exposure to allele X at a HLA locus with N alleles. The exposure to allele X was defined by the string with N-1 zeros and a one at position X.

The continuous variables that were used are weight and age of both recipient and donor and the ischemic time. For these variables, exposure was defined as the original value of the variable plus one. Weights were increased by one kg, ages were increased by one year and the ischemic time was increased by one hour.

For all discrete non-HLA variables, ORs were calculated relative to the most frequent value of the variable in the dataset. ORs of continuous variables were calculated relative to the original value of the variable. ORs of the HLA matching variables were calculated relative to no mismatches. The ORs of the alleles were calculated relative to the original HLA profile. In these calculations all transplants where the allele whose OR was to be calculated occurred were removed.

In the calculations of the ORs, the entire dataset was used. To calculate the odds of the variable values, the original values in the study population were temporarily replaced by the value whose odds was to be calculated. The probability for an event was evaluated for each transplant by the ensemble members. From the probabilities, the odds and the ORs for the variable values were obtained and the ORs for the individual transplants were averaged. The averaged OR of a variable value is called the effective OR [23].

To establish whether an allele was associated with high or low risk of early death, the ORs of the alleles were analyzed further by removing the 5th and 95th percentile and the OR calculations for an allele were considered unambiguous when both values lay below or above one.

3.7 Kaplan-Meier Survival Analysis

In a survival analysis the effect of an exposure on an event of interest in a study population is examined. Often the study population contains subjects that left the study before the event of interest had occurred. These subjects need to be handled in a survival analysis in order to gain maximum information from the data. A Kaplan-Meier (K-M) analysis is a way to perform a survival analysis on data that contains incomplete observations. The result of a K-M analysis can be presented as K-M curve [24].

To create a K-M curve, two variables are used for every subject in the study population. These are the time the patient participated in the study (serial time) and its status at the end of the serial time. As only the serial time is considered, all subjects can be analyzed starting from the same point in time. The time between the starting point and the desired end point of the survival analysis is then divided into intervals. An interval is defined by the time between two consecutive events of interest in the study population. A subject leaving the study for other reasons does not mark the end of an interval. For every interval the survival probability of the

study population is calculated. The survival probability of an interval that starts with a certain event is given by the ratio of the number of subjects that survive the event to the number of subjects that lived before the event. Subjects that leave the study during the interval are censored. In the calculation of the survival probability of the next interval, censored subjects are removed from the group of subjects that lived before the event [24].

In a K-M curve the cumulated probability of the intervals is shown. The cumulated probability of an interval is the product of the interval probabilities up to the event that starts the interval [24].

Often K-M curves are used to compare survival of two patient groups. The statistical significance of the difference in survival probabilities for the groups can be assessed in a Log-rank test [25].

In the K-M analysis, in this project, two transplant groups with 100 members each were compared. The high risk transplant group consisted of transplants where the allele with the highest effective OR whose 5th and 95th percentile were greater than one was involved. Transplants including this allele were sorted by the sum of effective ORs for all alleles involved and the 100 transplants with the highest sums were selected for the survival analysis. The low risk transplant group was made up of those transplants including the allele with smallest effective OR where both the 5th and 95th percentile were less than one. In this group, the 100 transplants with the smallest sum over the effective ORs of all involved alleles were selected for the survival analysis.

4 Results

In the first part of this section, the results of the data preprocessing and model calibration are presented. In the second part, the influence of HLA alleles on recipient short-term survival is studied. The performance of the network ensemble that was trained on the allele dataset is shown and the results of the sensitivity analysis and OR calculations are presented. This includes the identification of HLA alleles that could be associated with increased or decreased risk of early death unambiguously. The survival analysis of high- and low risk transplant groups is shown. In the third part of this section, the influence of HLA matching on recipient short-term survival is analyzed. The performance of the network ensemble that was trained on the matching dataset is shown and the results of the sensitivity analysis and OR calculations are presented.

4.1 Study Population

After removing incomplete transplant records, the final dataset consisted of 24,838 heart transplants. In the encoded allele dataset with detailed information about recipient and donor HLA alleles, every transplant was represented by a string of length 222. In the encoded HLA matching dataset, a transplant was represented by a string of length 41.

The HLA alleles that occurred in the final dataset were the same for recipients and donors. The number of HLA alleles at the study loci can be found in table 4.1. The occurrence of the HLA alleles in the study population can be found in table A.2 and A.3 for recipients and donors respectively. The frequency of non-HLA variable values can be found in table B.1. The frequency of the matching variables is shown in table B.2

Table 4.1: Number of alleles at the study loci in the final dataset. The alleles that occur in the final dataset are the same for recipients and donors.

Locus	Alleles
HLA-A	25
HLA-B	53
HLA-DR	17

4.2 Model Calibration

The optimal number of hidden layers and number of nodes per layer, the dropout p-values in the hidden and input layers and the regularization parameter λ were estimated using a cross-validated grid search. The corresponding parameters of the five best performing MLPs from the grid search are shown in table 4.2. The MLPs achieved an AUC of about 0.6 averaged over five validation folds.

The choice of the weight initialization, activation function for the hidden nodes and the optimizer was made based on fastest convergence of the AUC when a certain MLP architecture was trained. Convergence was achieved after 250 training epochs by a MLP with initial weights from a LeCun distribution, rectifier activation function in the hidden nodes and the Adam optimizer. The weight normalization parameter g was set to one.

Table 4.2: Parameters of the five best performing MLPs in the grid search.

Model	n hidden	n nodes	p hidden	p input	λ
MLP 1	1	10	0.4	0.2	0.001
MLP 2	2	8	0.3	0.1	0.001
MLP 3	1	12	0.4	0.2	0.001
MLP 4	1	12	0.3	0.3	0.001
MLP 5	1	10	0.3	0.1	0.01

4.3 Performance of Ensemble Members on Allele Dataset

The network ensemble was trained on the allele dataset. The validation performances of the five MLPs, averaged over five validation folds is shown in table 4.3.

Table 4.3: Validation performance of the five ensemble members averaged over five validation folds.

Model	Performance (AUC)
MLP 1	0.589 ± 0.006
MLP 2	0.595 ± 0.005
MLP 3	0.590 ± 0.006
MLP 4	0.590 ± 0.007
MLP 5	0.595 ± 0.005

4.4 Sensitivity of Variable Values in Allele Dataset

Sensitivities of non-HLA variable values are shown in figure 4.1. The variable values differed widely in their importance as predictors. The most important predictors were donor age (sens. $= 1.0 \pm 0.1$), ischemic time (sens. $= 0.5 \pm 0.3$) and an infection (sens. $= 0.5 \pm 0.2$). Recipient gender female was the least important predictor (sens. $= -0.01 \pm 0.03$). The sensitivities were associated with relatively large uncertainties. This indicates that ensemble members relied on the different variable values to different degrees. Detailed results of the sensitivity analysis for non-HLA variable values can be found in table A.1.

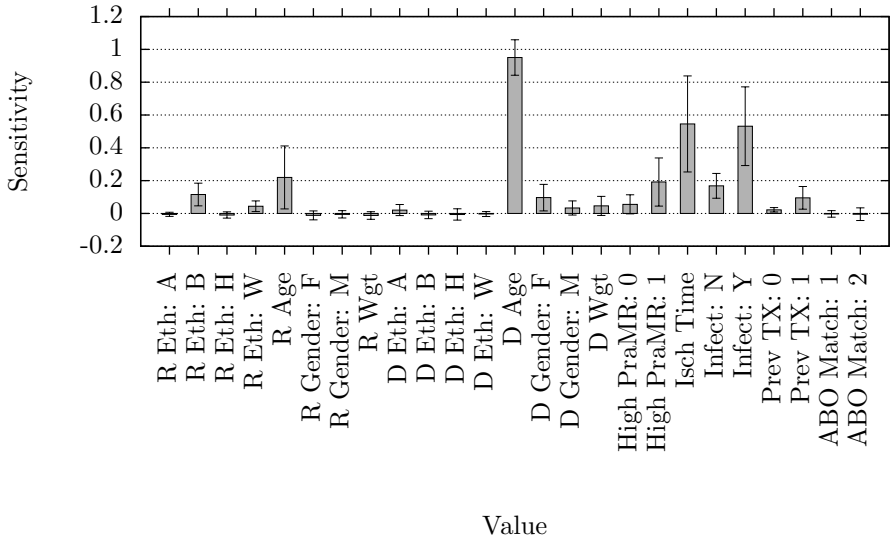


Figure 4.1: Sensitivity of non-HLA variable values for the allele dataset. The variable values differed widely in their importance as predictors. The uncertainty in the sensitivities shows that the importance of the predictors varied among the ensemble members.

In figure 4.2 the results of the sensitivity analysis for HLA alleles are shown in a box plot. In a box plot, a distribution of points is represented by a box whose lower and upper bound is given by the 25th percentile and the 75th percentile respectively. The mean is indicated by horizontal line across the box. The box is extended into both directions by whiskers that stretch to the most distant point that lies within 1.5 time the range between the 25th and 75th percentile. Points that are not captured by the whiskers are considered as outliers and are indicated by dots.

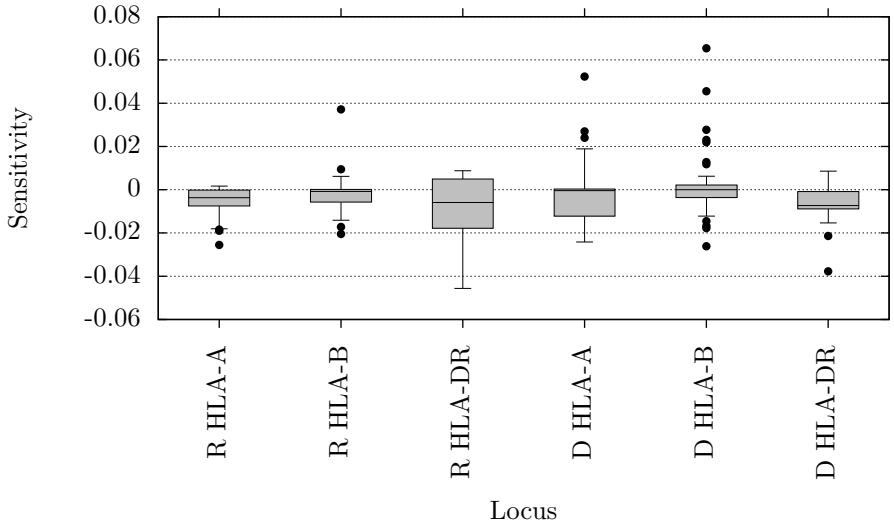


Figure 4.2: Box plot of sensitivities of HLA alleles at the study loci. The sensitivities at the loci fell mostly into narrow ranges below zero. Outliers with positive sensitivities could be observed at the recipients HLA-B locus and at the donors HLA-A and HLA-B loci.

In the box plot of the sensitivities, it can be seen that the ensemble has a negative average sensitivity to most HLA alleles. Sensitivities of the alleles at the study loci fell typically into a narrow ranges below zero. The largest spread in sensitivity could be observed at the recipients HLA-DR locus which was the only locus without outliers. At the recipients HLA-B and at the donors HLA-A and HLA-B loci, positive outliers could be observed. The locus with most outliers was the donors HLA-B locus. Detailed results of the sensitivity analysis for HLA alleles are shown in table A.2 and A.3 for the recipient and donor respectively.

The ten alleles with the highest sensitivity are shown in table 4.4. The ten alleles with highest sensitivity belonged to the HLA-A and -B loci. Only one allele was associated with the recipient (R HLA-B 35). The allele with highest sensitivity (D HLA-B 27) was the variable value with the 10th highest average sensitivity (sens. = 0.07 ± 0.05) which was comparable to the sensitivity of the value reduced immune activity (sens. = 0.06 ± 0.06).

Table 4.4: The ten alleles with highest sensitivity.

Ranking	Involvement	Locus	Allele
1	D	HLA-B	27
2	D	HLA-A	23
3	D	HLA-B	53
4	R	HLA-B	35
5	D	HLA-B	44
6	D	HLA-A	29
7	D	HLA-B	60
8	D	HLA-A	31
9	D	HLA-B	62
10	D	HLA-A	36

4.5 Odds Ratios of Variable Values in Allele Dataset

The effective ORs of non-HLA variables are shown in figure 4.3. The three most important risk factors for early death were an infection relative to no infection (eff. OR = 1.30), a high immune activity relative to a reduced immune activity (eff. OR = 1.18) and one previous transplant relative to no previous transplant (eff. OR = 1.15). Effective ORs for non-HLA variable values can be found in table A.1

The effective ORs of the alleles of the study loci are summarized in a box plot that is shown in figure 4.4. The alleles of the recipients HLA-B locus had the highest mean and it was the only locus where the 25th percentile was greater than one. The spread in the effective ORs among alleles at the same locus was larger compared to the sensitivity analysis and it was largest at the recipients HLA-DR locus which was the only locus without outliers with effective ORs less than one. Outliers with effective ORs greater than one could be observed at the recipient HLA-B locus and the donor HLA-A and HLA-B loci. The effective ORs for the HLA alleles can be found in table A.2 and A.3 for the recipient and donor respectively.

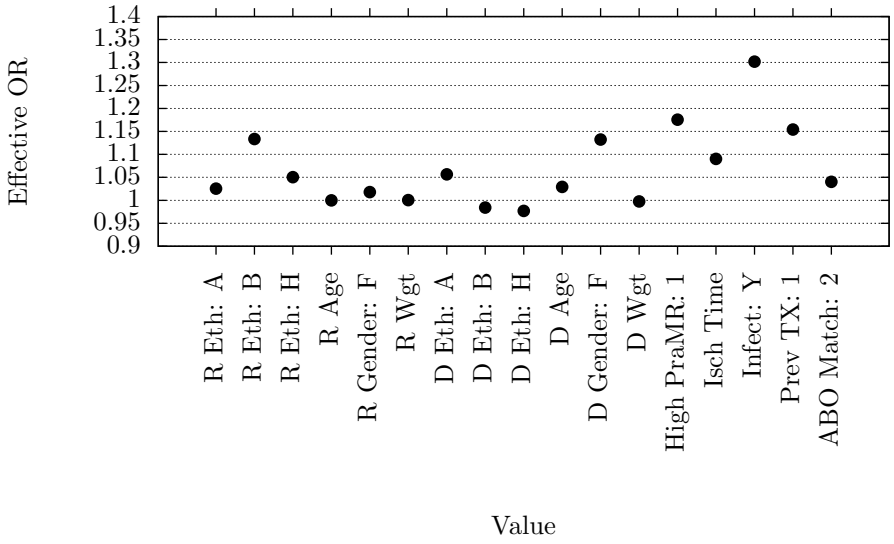


Figure 4.3: The effective ORs for non-HLA variable values. Most effective ORs fell into the range between 1 and 1.15. An infection relative to no infection is the most important risk factor.

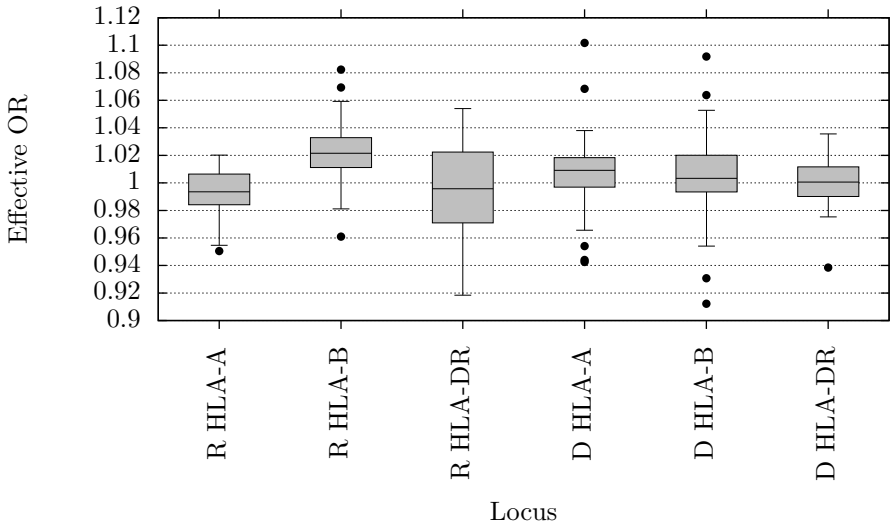


Figure 4.4: Box plot of the effective ORs of the HLA alleles at the study loci. Most effective ORs of the alleles range between 0.98 and 1.02. Outliers with effective ORs greater or less than one could be observed at all except for the recipients HLA-DR locus.

The 5th and 95th percentile of the ORs for individual alleles were calculated. For donor HLA-A 23 and HLA-B 53 alleles, the 5th and 95th percentile were greater than one while the corresponding values for the donor HLA-A 27 allele were less than one. The effective OR, the 5th an 95th percentile of the ORs and the occurrence in the study population for these alleles are shown in table 4.5.

Table 4.5: Effective OR, the 5th and 95 th percentile and occurrence of alleles with unambiguous result.

Involvement	Locus	Allele	Effective OR	5th perc.	95th perc.	Occ.
D	HLA-A	23	1.1017 ± 0.0009	1.0320	1.1933	1589
D	HLA-B	53	1.0918 ± 0.0009	1.0053	1.2096	951
D	HLA-B	27	0.9121 ± 0.0009	0.8334	0.9923	1791

4.6 Survival Curves

Groups of transplants with 100 members were compared in a K-M survival analysis. In high risk transplants, donors had the HLA-A 23 allele. In low risk transplants, donors had the HLA-B 27 allele. After one year, in the low risk group, 95 recipients were alive compared to 83 recipients in the high risk group. A significant difference in the survival probabilities of the transplant groups was found ($P = 0.0078$, Log-rank test). The K-M survival curves of the two groups are shown in figure 4.5.

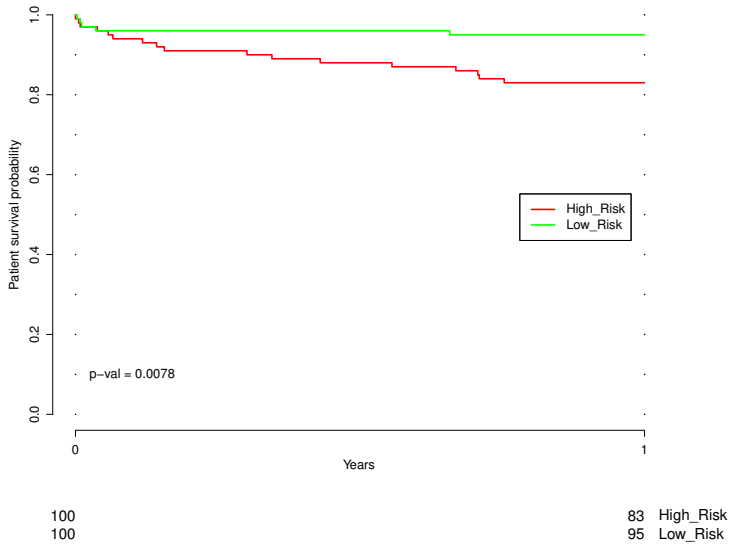


Figure 4.5: Survival curves of sorted high risk transplants with donor allele HLA-A 23 (red) and sorted low risk transplants with donor allele HLA-B 27 (green). The difference in curves was found to be significant ($P = 0.0078$, Log-rank test).

4.7 Performance of Ensemble Members on Matching Dataset

The validation performances of the ensemble members trained on the matching data averaged over the five validation folds is shown in table 4.6. The ensemble members had validation performances with an average AUC of about 0.6. This value was similar to the performances in the grid search where the allele dataset was used.

Table 4.6: Validation performance of the five ensemble members for the matching data averaged over five validation folds.

Model	Performance (AUC)
MLP 1	0.599 ± 0.006
MLP 2	0.600 ± 0.005
MLP 3	0.600 ± 0.007
MLP 4	0.599 ± 0.006
MLP 5	0.600 ± 0.005

4.8 Sensitivity of Variable Values in Matching Dataset

The sensitivity of non-HLA variable values is shown in figure 4.6. The importance of the variable values as predictors varied widely among the values. The three most important predictors were the donor age (sens = 0.9 ± 0.2), ischemic time (sens = 0.6 ± 0.3) and an infection (sens = 0.6 ± 0.2). Donor ethnicity Hispanic was the least important predictor (sens = -0.01 ± 0.03). The sensitivities were associated with relatively large uncertainties. Detailed results of the sensitivity analysis for non-HLA variable values can be found in table B.1.

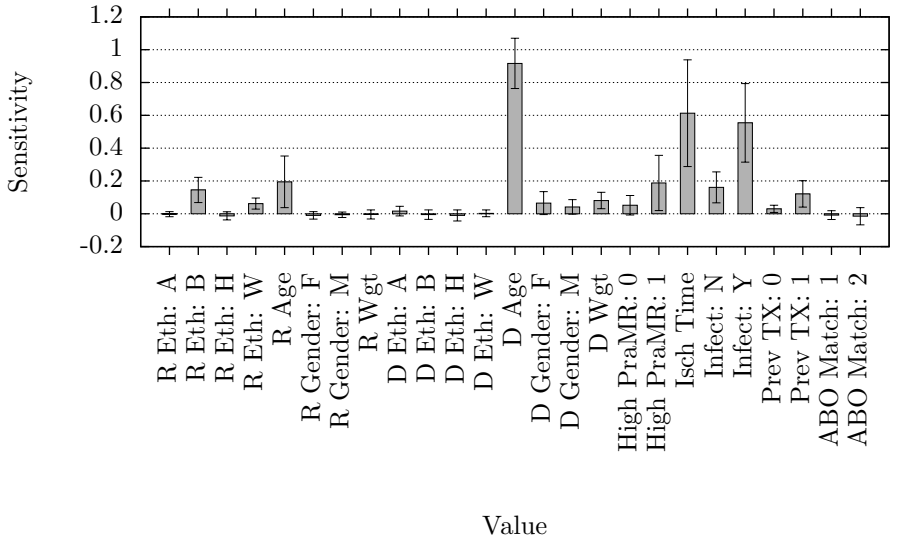


Figure 4.6: Sensitivity of non-HLA variable values in matching dataset. The variable values differed widely in their importance as predictors. The uncertainty in the sensitivities shows that the importance of the predictors varied among the ensemble members.

The sensitivity of HLA matching variable values is shown in figure 4.7. The matching variables were among the least important predictors and the only value with positive average sensitivity was one match at the HLA-B locus. The sensitivities were associated with large uncertainties. Sensitivities for the matching variables are shown in table B.2.

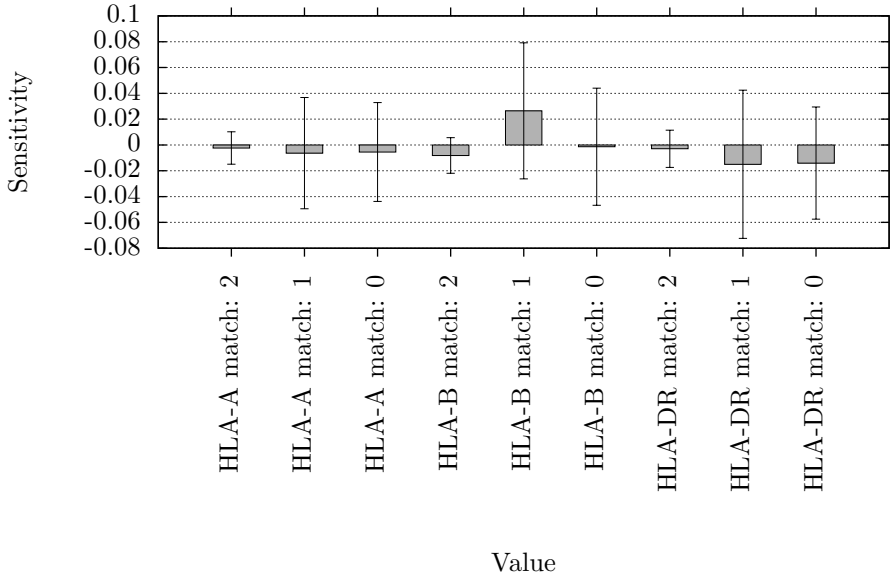


Figure 4.7: Sensitivity of HLA matching variable values. Only the value one mismatch at the HLA-B locus had a positive sensitivity. All values were associated with large uncertainties.

4.9 Odds Ratios of Variable Values in Matching Dataset

The effective ORs of the non-HLA variables are shown in figure 4.8. The most important risk factor was an infection relative to no infection (eff. OR = 1.41). Other important risk factors were one previous transplant relative to no previous transplant (eff. OR = 1.23) and an increased immune activity relative to a reduced immune activity (eff. OR = 1.23). The only value with an effective OR less than one was a Hispanic donor relative to a White donor (eff. OR = 0.98). The effective ORs for non-HLA variable values can be found in table B.1.

The effective ORs of HLA matching variable values, calculated relative to no mismatch, are shown in figure 4.9. For all loci, the value one mismatch was associated with an decreased risk of early death. At the HLA-A and -DR loci, two mismatches were associated with a decreased risk of early death. The only matching variable that was associated with an increased risk for early death was no match at the HLA-B locus (eff. OR = 1.03). The effective ORs for matching variables are shown in table B.2.

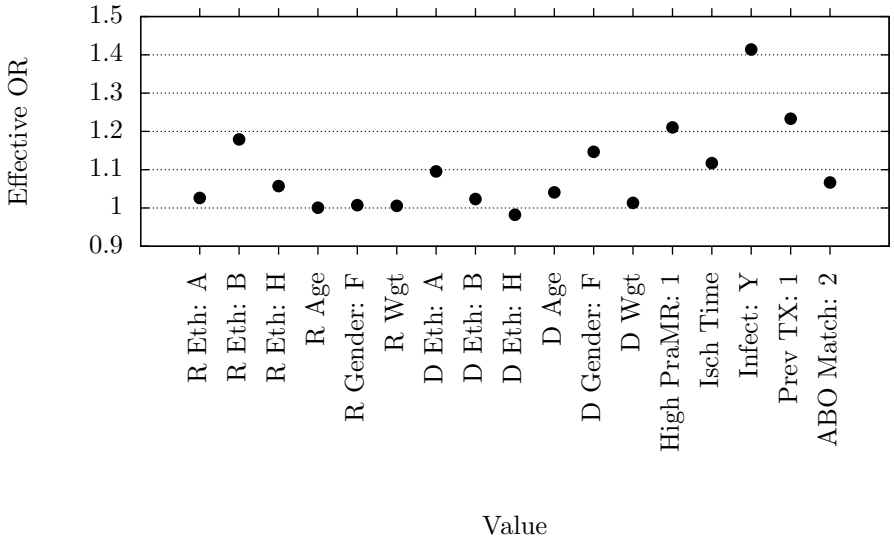


Figure 4.8: The effective ORs for non-HLA variable values. Most effective ORs fell into the range between 1 and 1.2. An infection relative to no infection was the most important risk factor.

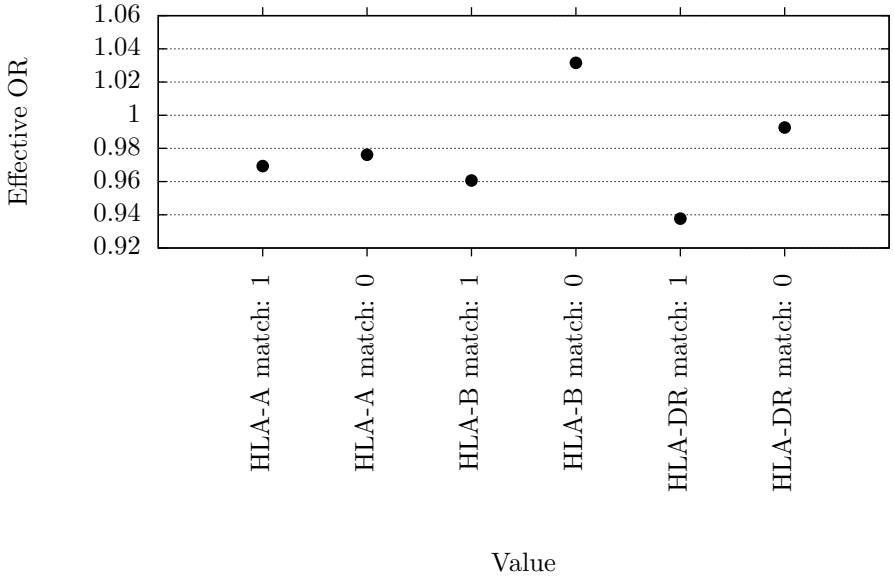


Figure 4.9: The effective ORs for HLA matching variables. The value one mismatch has an effective OR less than one at all loci. The only matching variable with effective OR greater than one is the value zero matches at the HLA-B locus.

5 Discussion

Records of more than twenty thousand heart transplants were used to evaluate the influence of HLA-A, -B and -DR alleles on recipient short-term survival in heart transplants using a network ensemble of five MLPs. A sensitivity analysis on the predictor variable values was performed and ORs were calculated to assess the HLA alleles as risk factors.

Unambiguous results according to the criterion used in this project could be established for three alleles. The donors HLA-A 23 and HLA-B 53 alleles were identified as risk factors for early death. The donors HLA-B 27 allele could be associated with decreased risk of early death. Among all alleles, these were the most important predictors. The results were validated in a K-M survival analysis where a significant difference in the survival probability of sorted high and low risk transplants with donor HLA-A 23 and donor HLA-B 27 alleles was found.

The effective ORs of the donors HLA-A 23 (eff. OR = 1.10) and HLA-B 53 (eff. OR = 1.09) alleles were the two highest of all alleles and comparable to the effective OR of an ischemic time that was increased by one hour (eff. OR = 1.09). The donors HLA-B 27 allele had the smallest effective OR of all alleles (eff. OR = 0.91). The donors HLA-A 23, HLA-B 53 and HLA-B 27 alleles were important predictors with sensitivities comparable to the variable value reduced immune activity (sens. = 0.05). Generally, in the sensitivity analysis and OR calculations it was shown that the donor HLA alleles were more important to recipient one-year survival than the recipient HLA alleles.

To our knowledge, this is a new result as no previous studies on the effect of individual HLA alleles on recipient short-term survival exist.

In the survival analysis, the difference in survival probabilities for two transplant groups including the HLA-A 23 and HLA-B 27 alleles respectively was found to be significant. In the selection of transplants for the two groups, even the effective ORs of the alleles at the other loci were considered. The difference in survival probabilities is thus not entirely due to the HLA-A 23 and HLA-B 27 alleles. The result gives still confidence that the OR calculations gave relevant results and that the HLA encoding procedure was appropriate.

To refine the result, the classifier performance needs to be improved. Shallow MLPs with AUCs less than 0.6 were used in the final data analysis. The architecture and parameters of the MLPs were partially determined in a grid search and AUCs of the MLPs that were used in the calculations were slightly smaller than in the

grid search. Better classification might, to some degree, be attained by refining the parameter values of the MLPs. MLPs with one and two hidden layers and between eight to twelve nodes in the hidden layers were used. Deeper MLPs had been tested but did not perform better. Network architectures different from MLPs using learning strategies other than supervised learning might achieve better classification results.

A more robust ensemble might be obtained by using more diverse MLPs in the network ensemble. The architectures and parameter values of the ensemble members were similar. Only one MLP had two hidden layers. The number of nodes in the hidden layers, the regularization parameters λ and dropout p-values were similar for all MLPs. Only the dropout p-value in the input layer spanned a wider range.

A larger AUC might possibly even be achieved by adding additional study variables. These should be linked to the HLA system.

In the sensitivity analysis HLA alleles were generally among the least important predictors. It could thus be concluded that most HLA alleles had a small impact on the classification results of the network ensemble. This could even be seen in an analysis that is not part of this report where the transplants were classified based on recipient and donor HLA alleles alone. These classifications had AUCs of about 0.5.

To improve the reliability of the result, more transplant records need to be included in the calculations. With respect to the UNOS database, this could be achieved by replacing missing values. However, replacing missing values bears the risk of introducing additional noise to the small signal from the HLA alleles. This noise could potentially outweigh the beneficial effect of an increased number of transplant records. If available, using a larger database should be considered. If there is no larger database available, the study result should be validated against external data to increase the confidence in the result.

The practical implications of the result are hard to assess. As there is a shortage of grafts, it cannot be afforded to waste donor hearts even if either of the two high risk alleles is present on the graft. An analysis of possible interactions of the high risk alleles and other alleles or variables could be performed. To do such an analysis, a larger dataset is required. Moreover, to assess the implication of the results, it is necessary to analyze the influence of the alleles on recipient long-term survival as well.

A smaller part of this work was dedicated to the analysis of the influence of HLA matching on recipient short-term survival. At all loci it has been found that one mismatch is associated with a smaller risk of early death compared to no mismatch. This applies even to two mismatches at the HLA-A and HLA-DR loci. This result does not agree with the naive theory presented in section 2 and previous studies that have established a negative effect of HLA mismatches on the recipients short-term survival [5].

6 Conclusions

The influence of HLA-A, -B and -DR alleles on recipient one-year survival in heart transplants was evaluated in a sensitivity analysis and by calculating ORs using an ensemble of five MLPs. The result of the calculations was validated in a K-M survival analysis. To study the influence of HLA matching, a sensitivity analysis and OR calculations were performed. In the study, 24,838 transplant records from the UNOS database were used and 19 risk variables including the two HLA alleles at the recipient and donor HLA-A, -B and -DR loci were analyzed.

In OR calculations, donor HLA-A 23 and HLA-B 53 alleles were unambiguously associated with an increased risk of early death and the donor HLA-B 27 allele was unambiguously associated with a decreased risk of early death. The three alleles were important predictors and a significant difference in survival probability for two transplant groups that contained donor HLA-A 23 and HLA-B 27 alleles respectively was found in the survival analysis. In calculation of ORs for HLA matching variables, at all loci, one mismatch was associated with a decreased risk of early death relative to no mismatch. At the HLA-A and -DR loci, two mismatches were associated with decreased risk of early death relative to no mismatch while two mismatches at the HLA-B locus were associated with an increased risk.

The network ensemble was successfully applied as binary classifier and the employed data encoding procedure has proven to be suitable. To improve the result further, a larger dataset should be used and the selection of ANNs for classification should be refined. A greater certainty in the result could be achieved by using other criteria in the selection of transplant groups for the K-M survival analysis and by validating the study result against external data. The result of the HLA matching analysis are in conflict with previous studies on this subject [5].

Acknowledgements

I would like to thank my supervisor Mattias Ohlsson for his expert advice on ANNs throughout this project, as well as Johan Nilsson for the discussions that helped in my work at the interdisciplinary frontiers of machine learning and medicine.

A Allele Dataset

Table A.1: Occurrence, sensitivity and effective OR for non-HLA variables on allele data.

Variable	Value	Occurrence	Sensitivity	Effective OR
R Eth	A(sian)	479 (1.9%)	-0.005 ± 0.01	1.0253 ± 0.0009
	H(ispanic)	1692 (6.8%)	-0.009 ± 0.02	1.0505 ± 0.0009
	B(lack)	3937 (15.9%)	0.12 ± 0.07	1.133 ± 0.001
	W(hite)	18730 (75.4%)	0.04 ± 0.03	1.000 ± 0.002
R Age		51.7 (12.7)	0.2 ± 0.2	0.9998 ± 0.0009
R Gender	F(emale)	5764 (23.2%)	-0.01 ± 0.03	1.018 ± 0.001
	M(ale)	19074 (76.8%)	-0.005 ± 0.02	1.000 ± 0.002
R Wgt		80.1 (17.0)	-0.01 ± 0.02	1.0005 ± 0.0008
D Eth	A(sian)	336 (1.4%)	0.02 ± 0.03	1.0566 ± 0.0009
	H(ispanic)	3443 (13.9%)	-0.006 ± 0.03	0.977 ± 0.001
	B(lack)	3302 (13.3%)	-0.009 ± 0.02	0.984 ± 0.001
	W(hite)	17757 (71.5%)	-0.004 ± 0.01	1.000 ± 0.002
D Age		31.6 (12.1)	1.0 ± 0.1	1.0292 ± 0.0009
D Gender	F(emale)	7302 (29.4%)	0.10 ± 0.08	1.132 ± 0.001
	M(ale)	17536 (70.6%)	0.03 ± 0.04	1.000 ± 0.001
D Wgt		79.9 (17.9)	0.05 ± 0.06	0.9976 ± 0.0008
High PraMR		3217 (13.0%)	0.2 ± 0.1	1.1757 ± 0.0009
Isch Time		3.1 (1.0)	0.5 ± 0.3	1.0901 ± 0.0009
Infect		2596 (10.5%)	0.5 ± 0.2	1.3020 ± 0.0009
Prev TX	No	23987 (96.6%)	0.02 ± 0.01	1.000 ± 0.005
	1 time	812 (3.3%)	0.10 ± 0.07	1.1540 ± 0.0009
	> 1 times	39 (0.1%)		
ABOMatch	1		-0.003 ± 0.02	1.000 ± 0.002
	2		-0.005 ± 0.04	1.040 ± 0.001

Table A.2: Occurrence, sensitivity and effective OR for recipient HLA-A, -B and -DR alleles.

Locus	Allele	Occurrence	Sensitivity	Effective OR
HLA-A	2	13781 (55.5%)	-0.02 ± 0.06	1.0098 ± 0.001
HLA-A	1	6321 (25.4%)	-0.008 ± 0.03	0.9926 ± 0.001
HLA-A	3	6236 (25.1%)	-0.02 ± 0.04	1.0064 ± 0.001
HLA-A	24	4254 (17.1%)	-0.008 ± 0.03	0.9877 ± 0.0009
HLA-A	11	2761 (11.1%)	-0.008 ± 0.02	0.9901 ± 0.0009
HLA-A	30	2125 (8.6%)	-0.008 ± 0.02	0.9901 ± 0.0009

Table A.2: Occurrence, sensitivity and effective OR for recipient HLA-A, -B and -DR alleles.

Locus	Allele	Occurrence	Sensitivity	Effective OR
HLA-A	23	1812 (7.3%)	-0.007 ± 0.05	1.0093 ± 0.0009
HLA-A	29	1785 (7.2%)	-0.003 ± 0.008	1.0026 ± 0.0009
HLA-A	68	1714 (6.9%)	-0.02 ± 0.04	1.0116 ± 0.0009
HLA-A	32	1570 (6.3%)	-0.006 ± 0.01	0.9844 ± 0.0009
HLA-A	26	1504 (6.1%)	-0.006 ± 0.03	0.9505 ± 0.0009
HLA-A	31	1233 (5.0%)	-0.01 ± 0.04	0.95838 ± 0.0009
HLA-A	33	1190 (4.8%)	-0.004 ± 0.01	1.0057 ± 0.0009
HLA-A	28	907 (3.7%)	0.0009 ± 0.008	0.9817 ± 0.0009
HLA-A	25	823 (3.3%)	-0.0004 ± 0.02	0.9546 ± 0.0009
HLA-A	74	497 (2.0%)	-0.0008 ± 0.006	0.9968 ± 0.0009
HLA-A	34	388 (1.6%)	0.002 ± 0.01	1.0202 ± 0.0009
HLA-A	66	322 (1.3%)	-0.001 ± 0.008	1.0110 ± 0.0009
HLA-A	36	248 (1.0%)	-0.004 ± 0.01	1.0105 ± 0.0009
HLA-A	69	75 (0.3%)	0.0003 ± 0.004	0.9754 ± 0.0009
HLA-A	80	64 (0.3%)	-0.0001 ± 0.003	0.9968 ± 0.0009
HLA-A	10	35 (0.1%)	-0.0003 ± 0.001	0.9859 ± 0.0009
HLA-A	19	19 (0.1%)	-0.0001 ± 0.002	0.9969 ± 0.0009
HLA-A	43	7 (0.0%)	-0.0001 ± 0.0004	0.9935 ± 0.0009
HLA-A	9	5 (0.0%)	$8e{-5} \pm 1e{-4}$	0.9841 ± 0.0009
HLA-B	44	6550 (26.4%)	-0.004 ± 0.05	1.000 ± 0.001
HLA-B	7	5569 (22.4%)	-0.02 ± 0.04	1.0054 ± 0.0009
HLA-B	35	4770 (19.2%)	0.04 ± 0.05	0.9609 ± 0.0009
HLA-B	8	4229 (17.0%)	-0.006 ± 0.04	0.9993 ± 0.0009
HLA-B	62	2595 (10.4%)	-0.01 ± 0.04	1.0050 ± 0.0009
HLA-B	51	2452 (9.9%)	-0.02 ± 0.03	1.0340 ± 0.0009
HLA-B	60	2243 (9.0%)	-0.01 ± 0.02	1.0123 ± 0.0009
HLA-B	18	2052 (8.3%)	-0.003 ± 0.02	0.9811 ± 0.0009
HLA-B	57	1923 (7.7%)	-0.005 ± 0.02	1.0480 ± 0.0009
HLA-B	27	1812 (7.3%)	-0.008 ± 0.03	1.0018 ± 0.0009
HLA-B	53	1249 (5.0%)	0.010 ± 0.009	1.0592 ± 0.0009
HLA-B	13	1053 (4.2%)	-0.002 ± 0.01	1.0409 ± 0.0009
HLA-B	39	1027 (4.1%)	-0.01 ± 0.03	1.0186 ± 0.0009
HLA-B	58	930 (3.7%)	-0.009 ± 0.01	1.0236 ± 0.0009
HLA-B	49	917 (3.7%)	-0.005 ± 0.07	1.0823 ± 0.0009
HLA-B	38	868 (3.5%)	-0.006 ± 0.02	1.0260 ± 0.0009
HLA-B	61	775 (3.1%)	0.0006 ± 0.01	1.0030 ± 0.0008
HLA-B	14	768 (3.1%)	-0.009 ± 0.02	1.0135 ± 0.0008
HLA-B	45	753 (3.0%)	-0.003 ± 0.007	1.0111 ± 0.0009
HLA-B	55	731 (2.9%)	-0.005 ± 0.01	1.0092 ± 0.0008
HLA-B	65	723 (2.9%)	-0.008 ± 0.01	1.0238 ± 0.0008
HLA-B	50	561 (2.3%)	0.0005 ± 0.008	1.0207 ± 0.0008
HLA-B	37	552 (2.2%)	0.001 ± 0.006	1.0358 ± 0.0008

Table A.2: Occurrence, sensitivity and effective OR for recipient HLA-A, -B and -DR alleles.

Locus	Allele	Occurrence	Sensitivity	Effective OR
HLA-B	52	546 (2.2%)	0.005 ± 0.01	1.0693 ± 0.0008
HLA-B	42	542 (2.2%)	-0.002 ± 0.02	1.0329 ± 0.0008
HLA-B	41	514 (2.1%)	-0.0008 ± 0.005	1.0380 ± 0.0008
HLA-B	70	400 (1.6%)	0.0009 ± 0.009	1.0512 ± 0.0008
HLA-B	63	390 (1.6%)	-0.0007 ± 0.008	1.0443 ± 0.0008
HLA-B	72	364 (1.5%)	-0.001 ± 0.01	1.0231 ± 0.0008
HLA-B	56	259 (1.0%)	-0.009 ± 0.02	1.0338 ± 0.0008
HLA-B	64	240 (1.0%)	-0.001 ± 0.007	1.0094 ± 0.0008
HLA-B	71	238 (1.0%)	0.001 ± 0.01	0.9979 ± 0.0008
HLA-B	81	161 (0.6%)	-0.002 ± 0.005	1.0293 ± 0.0008
HLA-B	15	146 (0.6%)	-0.005 ± 0.01	1.0209 ± 0.0008
HLA-B	40	136 (0.5%)	-0.004 ± 0.007	1.0247 ± 0.0008
HLA-B	48	111 (0.4%)	-0.0006 ± 0.003	1.0137 ± 0.0008
HLA-B	47	105 (0.4%)	-0.002 ± 0.004	1.0366 ± 0.0008
HLA-B	75	76 (0.3%)	0.001 ± 0.003	1.0016 ± 0.0008
HLA-B	78	67 (0.3%)	-0.0007 ± 0.004	1.0359 ± 0.0008
HLA-B	17	58 (0.2%)	-0.0004 ± 0.001	1.0207 ± 0.0008
HLA-B	46	54 (0.2%)	-0.0005 ± 0.002	1.0212 ± 0.0008
HLA-B	22	40 (0.2%)	0.0004 ± 0.001	1.0124 ± 0.0008
HLA-B	5	25 (0.1%)	0.0005 ± 0.001	1.0310 ± 0.0008
HLA-B	54	22 (0.1%)	0.0006 ± 0.0006	1.0097 ± 0.0008
HLA-B	73	17 (0.1%)	-0.0001 ± 0.0005	1.0216 ± 0.0008
HLA-B	82	15 (0.1%)	-0.0005 ± 0.001	1.0190 ± 0.0008
HLA-B	21	13 (0.1%)	$-3e{-}6 \pm 5e{-}4$	1.0145 ± 0.0008
HLA-B	16	12 (0.0%)	-0.0007 ± 0.001	1.0238 ± 0.0008
HLA-B	77	8 (0.0%)	0.0001 ± 0.0002	1.0151 ± 0.0008
HLA-B	67	5 (0.0%)	-0.00005 ± 0.0002	1.0239 ± 0.0008
HLA-B	12	4 (0.0%)	-0.0002 ± 0.0007	1.0268 ± 0.0008
HLA-B	76	3 (0.0%)	0.0001 ± 0.0006	1.0283 ± 0.0008
HLA-B	59	3 (0.0%)	-0.0001 ± 0.0003	1.0245 ± 0.0008
HLA-DR	4	7906 (31.8%)	0.006 ± 0.1	0.918 ± 0.001
HLA-DR	13	6446 (26.0%)	0.003 ± 0.1	1.053 ± 0.001
HLA-DR	15	6283 (25.3%)	-0.01 ± 0.05	1.022 ± 0.001
HLA-DR	7	6192 (24.9%)	-0.02 ± 0.05	1.000 ± 0.001
HLA-DR	11	5496 (22.1%)	0.003 ± 0.06	1.050 ± 0.001
HLA-DR	1	4547 (18.3%)	-0.04 ± 0.07	1.003 ± 0.001
HLA-DR	17	3458 (13.9%)	-0.02 ± 0.05	0.971 ± 0.001
HLA-DR	8	2009 (8.1%)	-0.005 ± 0.03	0.996 ± 0.001
HLA-DR	3	1629 (6.6%)	-0.02 ± 0.05	0.992 ± 0.001
HLA-DR	14	1422 (5.7%)	0.006 ± 0.05	1.054 ± 0.001
HLA-DR	12	1068 (4.3%)	-0.03 ± 0.04	0.985 ± 0.001
HLA-DR	16	789 (3.2%)	0.005 ± 0.02	0.964 ± 0.001

Table A.2: Occurrence, sensitivity and effective OR for recipient HLA-A, -B and -DR alleles.

Locus	Allele	Occurrence	Sensitivity	Effective OR
HLA-DR	9	780 (3.1%)	0.003 ± 0.02	0.964 ± 0.001
HLA-DR	10	572 (2.3%)	0.005 ± 0.02	1.041 ± 0.001
HLA-DR	2	569 (2.3%)	-0.007 ± 0.02	0.990 ± 0.001
HLA-DR	18	455 (1.8%)	0.006 ± 0.02	0.953 ± 0.001
HLA-DR	5	55 (0.2%)	0.0006 ± 0.002	1.008 ± 0.001

Table A.3: Occurrence, sensitivity and effective OR for donor HLA-A, -B and -DR alleles.

Locus	Allele	Occurrence	Sensitivity	Effective OR
HLA-A	2	13890 (55.9%)	-0.01 ± 0.05	1.029 ± 0.001
HLA-A	1	6506 (26.2%)	-0.01 ± 0.03	1.018 ± 0.001
HLA-A	3	6238 (25.1%)	-0.01 ± 0.05	0.999 ± 0.001
HLA-A	24	4278 (17.2%)	0.02 ± 0.06	0.9656 ± 0.0009
HLA-A	11	2641 (10.6%)	-0.006 ± 0.02	1.0181 ± 0.0009
HLA-A	30	2131 (8.6%)	-0.004 ± 0.02	1.0118 ± 0.0009
HLA-A	68	1957 (7.9%)	-0.01 ± 0.08	0.9541 ± 0.0009
HLA-A	29	1877 (7.6%)	0.03 ± 0.03	0.9440 ± 0.0009
HLA-A	23	1664 (6.7%)	0.05 ± 0.04	1.1017 ± 0.0009
HLA-A	32	1542 (6.2%)	-0.005 ± 0.01	1.0059 ± 0.0009
HLA-A	31	1456 (5.9%)	0.02 ± 0.02	0.9425 ± 0.0009
HLA-A	26	1320 (5.3%)	-0.02 ± 0.05	1.0184 ± 0.0009
HLA-A	33	1160 (4.7%)	-0.01 ± 0.03	0.9969 ± 0.0009
HLA-A	28	841 (3.4%)	-0.01 ± 0.04	1.0309 ± 0.0009
HLA-A	25	810 (3.3%)	-0.01 ± 0.04	1.0380 ± 0.0009
HLA-A	74	437 (1.8%)	-0.0004 ± 0.005	1.0250 ± 0.0009
HLA-A	34	320 (1.3%)	0.008 ± 0.009	0.9670 ± 0.0008
HLA-A	66	270 (1.1%)	-0.002 ± 0.01	1.0002 ± 0.0008
HLA-A	36	205 (0.8%)	0.02 ± 0.02	1.0683 ± 0.0008
HLA-A	69	53 (0.2%)	-0.0001 ± 0.004	0.9968 ± 0.0008
HLA-A	80	48 (0.2%)	0.0001 ± 0.002	1.0021 ± 0.0008
HLA-A	19	14 (0.1%)	0.0003 ± 0.0004	1.0024 ± 0.0008
HLA-A	10	12 (0.0%)	-0.0003 ± 0.0007	1.0108 ± 0.0008
HLA-A	9	4 (0.0%)	$3e{-6} \pm 8e{-5}$	1.0091 ± 0.0008
HLA-A	43	2 (0.0%)	$6e{-6} \pm 5e{-5}$	1.0139 ± 0.0008
HLA-B	44	6351 (25.6%)	0.02 ± 0.05	1.063 ± 0.001
HLA-B	7	5978 (24.1%)	0.006 ± 0.04	0.982 ± 0.001
HLA-B	35	4598 (18.5%)	-0.01 ± 0.03	1.0036 ± 0.0009
HLA-B	8	4436 (17.9%)	-0.01 ± 0.02	1.0262 ± 0.0009
HLA-B	62	2748 (11.1%)	0.02 ± 0.05	0.9541 ± 0.0009
HLA-B	51	2413 (9.7%)	-0.01 ± 0.04	1.0085 ± 0.0009
HLA-B	60	2211 (8.9%)	0.03 ± 0.05	0.9307 ± 0.0009

Table A.3: Occurrence, sensitivity and effective OR for donor HLA-A, -B and -DR alleles.

Locus	Allele	Occurrence	Sensitivity	Effective OR
HLA-B	18	1998 (8.0%)	-0.007 ± 0.02	1.0103 ± 0.0009
HLA-B	27	1838 (7.4%)	0.06 ± 0.05	0.9122 ± 0.0009
HLA-B	57	1757 (7.1%)	-0.006 ± 0.02	0.9903 ± 0.0009
HLA-B	39	1333 (5.4%)	-0.03 ± 0.03	0.9907 ± 0.0009
HLA-B	13	1060 (4.3%)	-0.005 ± 0.02	1.0045 ± 0.0009
HLA-B	53	989 (4.0%)	0.05 ± 0.04	1.0918 ± 0.0009
HLA-B	61	955 (3.8%)	0.004 ± 0.01	1.0298 ± 0.0009
HLA-B	58	944 (3.8%)	0.01 ± 0.02	1.0528 ± 0.0009
HLA-B	49	834 (3.4%)	-0.008 ± 0.02	1.0068 ± 0.0009
HLA-B	65	834 (3.4%)	0.005 ± 0.01	0.9589 ± 0.0009
HLA-B	14	749 (3.0%)	0.0008 ± 0.02	0.9920 ± 0.0009
HLA-B	38	744 (3.0%)	-0.01 ± 0.03	0.9935 ± 0.0009
HLA-B	55	725 (2.9%)	-0.006 ± 0.04	0.9585 ± 0.0009
HLA-B	45	714 (2.9%)	0.0008 ± 0.02	1.0292 ± 0.0009
HLA-B	52	560 (2.3%)	0.0002 ± 0.007	1.0212 ± 0.0009
HLA-B	50	547 (2.2%)	-0.003 ± 0.02	1.0258 ± 0.0009
HLA-B	37	542 (2.2%)	0.004 ± 0.02	1.0465 ± 0.0009
HLA-B	42	486 (2.0%)	-0.01 ± 0.03	0.9996 ± 0.0009
HLA-B	41	469 (1.9%)	-0.01 ± 0.03	1.0201 ± 0.0009
HLA-B	70	466 (1.9%)	0.004 ± 0.02	1.0480 ± 0.0009
HLA-B	72	353 (1.4%)	0.002 ± 0.01	0.9727 ± 0.0009
HLA-B	63	351 (1.4%)	-0.004 ± 0.01	0.9911 ± 0.0009
HLA-B	56	257 (1.0%)	0.01 ± 0.02	1.0474 ± 0.0009
HLA-B	64	256 (1.0%)	-0.004 ± 0.01	1.0209 ± 0.0009
HLA-B	71	222 (0.9%)	-0.001 ± 0.007	1.0100 ± 0.0009
HLA-B	48	208 (0.8%)	-0.002 ± 0.003	1.0032 ± 0.0008
HLA-B	81	132 (0.5%)	0.004 ± 0.006	0.9689 ± 0.0008
HLA-B	47	121 (0.5%)	-0.002 ± 0.004	0.9988 ± 0.0008
HLA-B	40	96 (0.4%)	0.0005 ± 0.005	1.0185 ± 0.0008
HLA-B	75	91 (0.4%)	0.0009 ± 0.003	0.9924 ± 0.0008
HLA-B	15	52 (0.2%)	-0.0005 ± 0.001	1.0033 ± 0.0008
HLA-B	78	48 (0.2%)	-0.0007 ± 0.002	1.0013 ± 0.0008
HLA-B	17	46 (0.2%)	0.002 ± 0.003	1.0251 ± 0.0008
HLA-B	46	44 (0.2%)	-0.0002 ± 0.001	1.0060 ± 0.0008
HLA-B	73	19 (0.1%)	$2e{-5} \pm 5e{-4}$	0.9987 ± 0.0008
HLA-B	54	19 (0.1%)	-0.0002 ± 0.0007	1.0064 ± 0.0008
HLA-B	5	15 (0.1%)	-0.0005 ± 0.0009	1.0030 ± 0.0008
HLA-B	22	13 (0.1%)	$8e{-5} \pm 4e{-4}$	1.0005 ± 0.0008
HLA-B	82	12 (0.0%)	$2e{-5} \pm 9e{-4}$	1.0060 ± 0.0008
HLA-B	16	11 (0.0%)	-0.0001 ± 0.0003	1.0046 ± 0.0008
HLA-B	59	8 (0.0%)	$8e{-5} \pm 1e{-4}$	1.0000 ± 0.0008
HLA-B	21	7 (0.0%)	$-2e{-5} \pm 8e{-4}$	1.0105 ± 0.0008

Table A.3: Occurrence, sensitivity and effective OR for donor HLA-A, -B and -DR alleles.

Locus	Allele	Occurrence	Sensitivity	Effective OR
HLA-B	67	5 (0.0%)	$4e{-5} \pm 9e{-5}$	1.0008 ± 0.0008
HLA-B	77	5 (0.0%)	$2e{-5} \pm 9e{-5}$	1.0000 ± 0.0008
HLA-B	12	4 (0.0%)	$1e{-5} \pm 6e{-5}$	1.0011 ± 0.0008
HLA-B	76	2 (0.0%)	$8e{-6} \pm 3e{-5}$	1.0033 ± 0.0008
HLA-DR	4	8236 (33.2%)	-0.009 ± 0.01	0.996 ± 0.001
HLA-DR	15	6315 (25.4%)	-0.02 ± 0.03	0.996 ± 0.001
HLA-DR	13	6226 (25.1%)	0.003 ± 0.04	1.036 ± 0.001
HLA-DR	7	6110 (24.6%)	-0.006 ± 0.02	1.008 ± 0.001
HLA-DR	1	4616 (18.6%)	-0.008 ± 0.02	1.012 ± 0.001
HLA-DR	11	4591 (18.5%)	-0.004 ± 0.03	0.975 ± 0.001
HLA-DR	17	3974 (16.0%)	-0.01 ± 0.04	0.984 ± 0.001
HLA-DR	8	2319 (9.3%)	-0.04 ± 0.09	0.9776 ± 0.0009
HLA-DR	14	1622 (6.5%)	0.008 ± 0.03	0.9384 ± 0.0009
HLA-DR	3	1451 (5.8%)	-0.01 ± 0.03	1.0093 ± 0.0009
HLA-DR	12	989 (4.0%)	-0.0007 ± 0.01	0.9920 ± 0.0009
HLA-DR	16	743 (3.0%)	-0.007 ± 0.02	1.0027 ± 0.0009
HLA-DR	9	740 (3.0%)	-0.008 ± 0.01	1.0161 ± 0.0009
HLA-DR	2	724 (2.9%)	-0.0008 ± 0.03	1.0303 ± 0.0009
HLA-DR	10	556 (2.2%)	0.004 ± 0.01	1.0266 ± 0.0009
HLA-DR	18	435 (1.8%)	-0.004 ± 0.02	1.0006 ± 0.0009
HLA-DR	5	29 (0.1%)	$2e{-5} \pm 1e{-3}$	0.9902 ± 0.0009

B Matching Dataset

Table B.1: Occurrence, sensitivity and effective OR for non-HLA variables on matching data.

Variable	Value	Occurrence	Sensitivity	Effective OR
R Eth	A(sian)	479 (1.9%)	-0.003 ± 0.01	1.0261 ± 0.0006
	H(ispanic)	1692 (6.8%)	-0.01 ± 0.02	1.0573 ± 0.0006
	B(lack)	3937 (15.9%)	0.15 ± 0.08	1.1791 ± 0.0006
	W(hite)	18730 (75.4%)	0.06 ± 0.03	1.0000 ± 0.0006
R Age		51.7 (12.7)	0.2 ± 0.2	1.0009 ± 0.0006
R Gender	F(emale)	5764 (23.2%)	-0.009 ± 0.02	1.0074 ± 0.0006
	M(ale)	19074 (76.8%)	-0.006 ± 0.02	1.0000 ± 0.0006
R Wgt		80.1 (17.0)	-0.004 ± 0.03	1.0055 ± 0.0006
D Eth	A(sian)	336 (1.4%)	0.02 ± 0.03	1.0956 ± 0.0006
	H(ispanic)	3443 (13.9%)	-0.01 ± 0.03	0.9822 ± 0.0006
	B(lack)	3302 (13.3%)	-0.005 ± 0.03	1.0235 ± 0.0006
	W(hite)	17757 (71.5%)	0.003 ± 0.02	1.0000 ± 0.0006
D Age		31.6 (12.1)	0.9 ± 0.2	1.0406 ± 0.0006
D Gender	F(emale)	7302 (29.4%)	0.06 ± 0.07	1.1466 ± 0.0006
	M(ale)	17536 (70.6%)	0.04 ± 0.04	1.0000 ± 0.0006
D Wgt		79.9 (17.9)	0.08 ± 0.05	1.0133 ± 0.0006
High PraMR		3217 (13.0%)	0.2 ± 0.2	1.2104 ± 0.0006
Isch Time		3.1 (1.0)	0.6 ± 0.3	1.1172 ± 0.0006
Infect		2596 (10.5%)	0.6 ± 0.2	1.4140 ± 0.0007
Prev TX	No	23987 (96.6%)	0.03 ± 0.02	1.0000 ± 0.0007
	1 time	812 (3.3%)	0.12 ± 0.08	1.2330 ± 0.0007
	> 1 times	39 (0.1%)		
ABO Match	1		-0.008 ± 0.03	1.0000 ± 0.0006
	2		-0.01 ± 0.05	1.0665 ± 0.0006

Table B.2: Occurrence, sensitivity and effective OR for HLA matching variables.

Locus	Value	Occurrence	Sensitivity	Effective OR
HLA-A	2 matches	1919 (7.7%)	-0.002 ± 0.01	1.0000 ± 0.0006
	1 matches	8062 (32.5%)	-0.006 ± 0.04	0.9693 ± 0.0005
	0 matches	14857 (59.8%)	-0.005 ± 0.04	0.9761 ± 0.0006
HLA-B	2 matches	558 (2.2%)	-0.008 ± 0.01	1.0000 ± 0.0006
	1 matches	5096 (20.5%)	0.03 ± 0.05	0.9607 ± 0.0006
	0 matches	19184 (77.2%)	-0.001 ± 0.05	1.0316 ± 0.0006
HLA-DR	2 matches	1317 (5.3%)	-0.003 ± 0.01	1.0000 ± 0.0006
	1 matches	7389 (29.7%)	-0.01 ± 0.06	0.9377 ± 0.0006
	0 matches	16132 (64.9%)	-0.01 ± 0.04	0.9925 ± 0.0006

References

[1] NIH. Heart transplant. `https://www.nhlbi.nih.gov/health/health-topics/topics/ht`, December 2016.

[2] Sharon A Hunt and François Haddad. The changing face of heart transplantation. *Journal of the American College of Cardiology*, 52(8):587–598, 2008.

[3] Sandrigo Mangini, Bárbara Rubim Alves, Odílson Marcos Silvestre, Philippe Vieira Pires, Lucas José Tachotti Pires, Milena Novaes Cardoso Curiati, and Fernando Bacal. Heart transplantation: review. *Einstein (São Paulo)*, 13(2):310–318, 2015.

[4] Marco Antonio Ayala García, Beatriz González Yebra, Andrea Liliana López Flores, and Eduardo Guaní Guerra. The major histocompatibility complex in transplantation. *Journal of transplantation*, 2012, 2012.

[5] David Ansari, Dragan Bućin, and Johan Nilsson. Human leukocyte antigen matching in heart transplantation: systematic review and meta-analysis. *Transplant International*, 27(8):793–804, 2014.

[6] María Pérez-Ortiz, Silvia Jiménez-Fernández, Pedro A Gutiérrez, Enrique Alexandre, César Hervás-Martínez, and Sancho Salcedo-Sanz. A review of classification problems and algorithms in renewable energy applications. *Energies*, 9(8):607, 2016.

[7] Mattias Ohlsson. Lecture notes on artificial neural networks (fytn06). 11 2015.

[8] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, volume 30, 2013.

[9] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[10] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

[11] Jacqueline Parkin and Bryony Cohen. An overview of the immune system. *The Lancet*, 357(9270):1777–1789, 2001.

[12] Francisco A Bonilla and Hans C Oettgen. Adaptive immunity. *Journal of Allergy and Clinical Immunology*, 125(2):S33–S40, 2010.

[13] Elizabeth Ingulli. Mechanism of cellular rejection in transplantation. *Pediatric nephrology*, 25(1):61, 2010.

[14] WM Howell, V Carter, and B Clark. The hla system: immunobiology, hla typing, antibody screening and crossmatching techniques. *Journal of clinical pathology*, 63(5):387–390, 2010.

[15] Steven GE Marsh, Peter Parham, and Linda D Barber. *The HLA factsbook*. Academic Press, 1999.

[16] LH Lund, LB Edwards, AY Kucheryavaya, AI Dipchand, C Benden, JD Christie, F Dobbels, R Kirk, AO Rahmel, RD Yusen, et al. International society for heart and lung transplantation. the registry of the international society for heart and lung transplantation: thirtieth official adult heart transplant report–2013; focus theme: age. *J Heart Lung Transplant*, 32(10):951–964, 2013.

[17] UNOS. Data. `https://www.unos.org/data/`, May 2017.

[18] Mark H Zweig and Gregory Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, 39(4):561–577, 1993.

[19] Xiao-Hua Zhou, Donna K McClish, and Nancy A Obuchowski. *Statistical methods in diagnostic medicine*, volume 569. John Wiley & Sons, 2009.

[20] François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.

[21] Warren S. Sarle. How to measure importance of inputs? `ftp://ftp.sas.com/pub/neural/importance.html`, 2000.

[22] Magdalena Szumilas. Explaining odds ratios. *Journal of the Canadian academy of child and adolescent psychiatry*, 19(3):227, 2010.

[23] Richard P Lippmann and David M Shahian. Coronary artery bypass risk prediction using neural networks. *The Annals of thoracic surgery*, 63(6):1635–1643, 1997.

[24] Jason T Rich, J Gail Neely, Randal C Paniello, Courtney CJ Voelker, Brian Nussenbaum, and Eric W Wang. A practical guide to understanding kaplan-meier curves. *Otolaryngology-Head and Neck Surgery*, 143(3):331–336, 2010.

[25] J Martin Bland and Douglas G Altman. The logrank test. *Bmj*, 328(7447):1073, 2004.