On the prediction of Hodgkin Lymphoma treatment response

Enrique J. deAndrés-Galiana [1,2,*], Juan Luis Fernández-Martínez [1,*], Oscar Luaces [2], Juan Jose del Coz [2], Rubén Fernández [3], Jackeline Solano [4], Eva Alonso Nogués [4], Yahya Zanabilli [5], Jose Manuel Alonso [6], Angel R. Payer [4], Jose María Vicente [7], Jesús Medina [5], Francisco Taboada [8], Manuel Vargas [9], Clara Alarcón [5], Marlen Morán [5], Angel González-Ordóñez [5], Miguel Angel Palicio [9], Sergio Ortiz [9], Cristina Chamorro [10], Segundo Gonzalez [11] and Ana P. González-Rodríguez [4]

[1] Department of Mathematics, University of Oviedo, Spain.
[2] Artificial Intelligence Center, University of Oviedo, Spain.
[3] Hematology Department, Hospital de Cabueñes, Gijon, Spain.
[4] Hematology Department, Hospital Universitario Central de Asturias, Oviedo, Spain.
[5] Hematology Department, Hospital San Agustin, Aviles, Spain.
[6] Hematology Department, Hospital Valle del Nalon, Langreo, Spain.
[7] Hematology Department, Hospital de Mieres, Mieres, Spain.
[8] Hematology Department, Hospital Cangas de Narcea, Cangas de Narcea, Spain.
[9] Hematology Department, Hospital de Jarrio, Jarrio, Spain.
[10] Hematology Department, Hospital de Arriondas, Arriondas, Spain.
[11] Instituto Universitario Oncológico del Principado de Asturias (IUOPA), University of Oviedo, Spain.
[*] Both authors have contributed equally to this study.

Corresponding author: Juan Luis Fernández-Martínez, jlfm@uniovi.es, 0034 985 103 199.

Abstract

Purpose: The cure rate in Hodgkin Lymphoma is high, but the response along the treatment is still unpredictable and is highly variable among patients. Detecting those patients that do not respond to the treatment at early stages could bring improvements in their treatment. This research tries to identify the main biological prognostic variables currently gathered at diagnosis, and designing a simple machine learning methodology to help physicians improving the treatment response assessment. **Methods**: We carried out a retrospective analysis of the response to treatment for a cohort of 263 Caucasians who were diagnosed with Hodgkin Lymphoma in Asturias (Spain). For that purpose, we used a list of 35 clinical and biological variables that are currently measured at diagnosis, before any treatment begins. To establish the list of most discriminatory prognostic variables for the treatment response we designed a machine learning approach based on two different feature selection methods (Fisher's ratio and Maximum Percentile Distance) and recursive feature elimination using a nearest-neighbor classifier (k-NN). The weights of the k-NN classifier are optimized using different terms of the confusion matrix (true and false positive rates) in order to minimize risk in the decisions. **Results and conclusions**: We found that the optimum strategy to predict treatment response in Hodgkin lymphoma consists in solving two

different binary classification problems, discriminating first if the patient is in progressive disease, if not, then discerning among complete and partial remission. Serum Ferritin turned to be the most discriminatory variable in predicting treatment response, followed by Alanine Aminotransferase and Alkaline Phosphate. The importance of these prognostic variables suggests a close relationship between inflammation, iron overload, liver damage and the extension of the disease.
Keywords: Hodgkin Lymphoma, treatment response, machine learning, Serum Ferritin (SF), Alanine Aminotransferase (ALT), Alkaline Phosphate (ALP).

PURPOSE

Lymphoma is the most common blood cancer and comprises two types: Hodgkin Lymphoma (HL) and non-Hodgkin Lymphoma. HL is characterized by the presence of the so-called malignant Reed-Sternberg cells, surrounded by an inflammatory infiltrate consisting of lymphocytes, neutrophils, eosinophils, plasma cells, macrophages and fibroblasts, constituting a model of interaction of tumor cells with their microenvironment. Components of inflammatory background are associated with classical HL: the presence of tumor-infiltrating lymphocytes is a negative prognostic factor for survival in these patients [1]. This kind of cancer is most commonly diagnosed in young adults between the ages of 15 and 35 and in older adults over 50. The cure rate in HL patients is high, but the response along the treatment is still unpredictable and varies from patient to patient. Besides, small minority is resistant or relapses before treatment. Detecting those patients with a poor prognosis at early stages (diagnosis) could bring improvements in their treatment and prognosis.

There was an international effort to identify prognostic factors to accurately predict the development and treatment in HL, mainly in patients with advanced HL stage. The adverse prognostic factors identified were: male older than 45, stage IV disease, hemoglobin lower than 10.5 g/dl, lymphocyte count lower than 600 /µl (or less than 8%), albumin lower than 4.0 g/dl, and white blood count greater than 15,000 /µl [2,

3]. Other studies also took into account mixed-cellularity or lymphocyte-depleted histologies, the presence of B symptoms or high erythrocyte sedimentation rate, and bulky disease as adverse prognostic factors [4, 5]. Moreover, disease extensions measured by Computed Tomography (CT), and early response to treatment measured by Positron Emission Tomography (PET) have demonstrated a powerful prognostic ability [6, 7].

Several research works highlighted the importance of the identification of prognostic variables to predict patients who will suffer relapse and the adaptation of treatments to individual risks [8-11]. Particularly, the result of the treatment optimization provoked some criteria modification, with the disappearance of some factors that were considered of poor prognosis and with the proposal of new ones that allowed establishing groups with differing risks of relapse and differentiated treatment.

In this manuscript we inferred prognostic variables for HL treatment response using clinical data and machine learning techniques in a retrospective study for a cohort of 263 Caucasians. For this purpose we designed a methodology to find the shortest list of clinical variables providing the highest predictive accuracy for the Hodgkin Lymphoma first line treatment response (at diagnosis). We found that the best way of addressing this problem is to proceed in two steps: comparing first the complete/partial remission hypothesis against progressive disease hypothesis, and secondly differentiating between complete and partial remission in case it proceeds. Serum Ferritin (SF) turned to be the most important prognostic variable, achieving cross-validation predictive accuracies higher than 90%. Ferritin concentrations increase drastically in the presence of an infection or cancer [12]. Our study has also showed the importance of the Alanine Aminotransferase (ALT) and the Alkaline Phosphate (ALP). The normal ranges for these three prognostic factors are provided in table S1 (see

Supplementary Material). The importance of these variables in the treatment response suggests a close relationship to iron overload, liver damage and bone affectation. An adequate staging of newly diagnosed patients using this methodology will enable optimal treatment planning, which is particularly important in health care in order to find an optimum balance between treatment efficacy and drug toxicity.

METHODS

The present research work is a retrospective study in a cohort of 263 Caucasians who were diagnosed with classical Hodgkin Lymphoma in Asturias (Spain) that were enrolled in this study between 2002 and 2012. This study was approved by the institutional review boards of the different hospitals involved, and it was performed in accordance with the Helsinki Declaration of 1975. Besides, this study was approved by the Ethics Committee of the Principado de Asturias (date: 17th of January; Project 6th number 13).

Staging definitions from the German Hodgkin Study Group (GHSG) were evaluated in this analysis. All patients were treated with the ABVD (doxorubicin, bleomycin, vinblastine, dacarbazine) regimen: 125 patients (47%) received involved field radiotherapy, 91 with early stage and 34 with advanced stage disease. Response to therapy was evaluated by physical and radiographic evaluation, including Computed Tomography (CT), and the follow up of the patients. In the last five years, PET scan was also included to assess treatment response. The treatment response was divided into 3 categories according to international standards [13]: 237 of the patients were in Complete Remission (CR), 17 in Partial Remission (PR): and only in 9 of the cases, the disease progressed without any relevant change. This last category was named as Progressive Disease (PD). Table 1 describes the main characteristics of the patients: age, sex, stage at diagnosis, percentage of early favorable and early unfavorable and

percentage of advanced disease depending on Hasenclever-Prognostic score.

Progression-Free Survival (PFS) was calculated from the date of diagnosis to the date of progression, relapse or death by of any cause. Overall Survival (OS) was calculated from the date of diagnosis to the date of death from any cause or last follow-up. Overall and progression-free survival distribution curves were estimated using the product-limit method of Kaplan–Meier. The median PFS and OS for the entire group were respectively 150 and 160 months. The probabilities of PFS and OS at 7 years were 57 and 76 %, correspondingly.

Thirty-five clinical and biological variables were measured at diagnosis and before treatment. These variables were classified into 5 groups: biochemical, inmuno-histochemical, Hodgkin Lymphoma specific, treatment specific, and host information. Table 2 shows the description of all these variables, boldfacing those that take discrete predefined values. Most of the variables had a sampling frequency higher than 90%. However, others were scarcely sampled, such as CRP (14%), immunoglobulins and Ki67 (20%). The need of imputing/filtering those variables has turned out to be a very important step in the modeling process.

The problem addressed in this manuscript consists in building an efficient and simple machine methodology to predict HL first line treatment response with the highest predictive accuracy, and at the same time minimizing risk in the decisions. For that purpose we have used the response criteria defined by Cheson et al (2007) [13]. Patients were divided into three categories: Complete Remission (CR), defined as the disappearance of all evidence of disease; Partial Remission (PR) defined as regression of measurable disease and no new sites; and Progressive Disease (PD) defined as any new lesion or increase by 50% of previously involved sites. Four different classification problems were performed, to find the optimum way of separating these three classes.

The machine learning methodology is explained in Appendix 1 (see Supplementary Material) and is composed by three main steps: 1. Pre-processing. 2. Feature Selection. 3. k-NN design and risk analysis.

Basically, the learning method consists, according to the parsimony principle, in finding the shortest subset of most discriminatory clinical variables (also called the reduced base of prognostic variables) to predict treatment response in HL patients. The clinical and biological variables are first ranked according to two different filter methods: Fisher's ratio (FR) and Maximum Percentile Distance (MPD). In a second step, the predictive accuracy of the different ranked lists of prognostic variables is established by Leave-One-Out-Cross-Validation (LOOCV) experiment using a simple k-Nearest-Neighbor (k-NN) classifier (Appendix 1 in Supplementary Material). This methodology has been successfully applied to predict risk of radiotherapy-related fatigue in prostate cancer patients using high dimensional expression data [14]. In this case, the challenge is not related to the dimension of the dataset, but to the heterogeneous degree of sampling of the different clinical variables. Besides, in this case the methodology incorporates the weights optimization of the k-NN classifier according to the Receiver Operating Characteristic (ROC) curve, in order to improve risk decision-making, that is, to provide a very high predictive accuracy with an optimum balance between the different rates of the confusion matrix (the true positive and false positive rates defining the corresponding ROC curve). Figure 1 shows a flow diagram explaining the methodology.

Finally, we would like to point out that the aim of this work is not to numerically comparing different machine learning methods, but to introduce simple algorithms to select the shortest list of prognostic variables that could be easily interpreted by medical doctors, in order to improve the patient prognostic in HL treatment response, with its

corresponding risk assessment. Particularly, we tried to avoid the use of black boxes that provide estimations without Medical Doctors' understanding. As a matter of fact, this methodology can be easily implemented in any platform such as a spreadsheet (see Supplementary Material –HLtreatementResponse_Predictor.xls-, and the corresponding explanation provided in Appendix 2). That said, the classifier that is proposed in this paper outperformed other more sophisticated classifiers that are proposed in the machine learning literature, highlighting the importance of selecting the correct prognostic variables.

RESULTS

Treatment response in HL is a difficult prediction problem. Aside from plasma EBV DNA [15], there is no predictive biomarker to bode the patient's response to the corresponding treatment with a reliable accuracy. This classification problem is intrinsically highly unbalanced, mainly due to the discrete sampling of the samples (number of patients), and also due to the fact that a high percentage of the patients are cured from this kind of malignancy.

The first modeling decision was to transform the analysis of treatment response into a binary classification problem (two-class problem) that admits a more reliable and stable solution than the corresponding value regression problem, that is, it is easier to predict if a patient is in complete or partial remission than predicting the value of the biological variables related to this fact. Besides, the prediction in binary classification problems allows for risk assessment through the analysis of the confusion matrix and the Receiving Operating Characteristic (ROC) curve. The confusion matrix consists of 4 different groups: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN); whose definition depends on how the classification problem has been set up. From the confusion matrix, different rates can be calculated to understand

the risk on the prediction:

1. True Positive Rate or Sensitivity (TPR): measures the proportion of actual positives that are correctly predicted as such.
2. True Negative Rate or Specificity (SPC): measures the proportion of negatives that are correctly predicted as such.
3. False Positive Rate (FPR): fraction of false positives out of the total actual negatives.
4. False Negative Rate (FNR): fraction of false negatives out of the total actual positives.
5. False Discovery Rate (FDR): fraction of false positives out of the total actual positives.

These rates could be used by the physicians in their decision-making process. A perfect classifier would have 100% sensitivity and specificity.

The following comparisons were performed:

1. CR vs. PR+PD
2. CR+PR vs. PD and CR vs. PR
3. CR vs. PR vs. PD

Comments for the prognostic variables in comparisons 1 and 3 are given in Appendix 3 (see Supplementary Material), since we have obtained worse results. The most effective comparison was the second one, and it is composed of two main steps. In the first step (2.1 CR+ PR vs. PD), we established the differences between patients that experimented partial or complete remission (CR + PR, positive class) from those in which the disease progressed without any relevant change (PD, negative class). Then, a second comparison (2.2 CR vs. PR) is used to establish the differences between CR (positive class) and PR (negative class) patients.

The best result was obtained by filtering out those variables having a sampling frequency lower than 30%, and imputing the rest. Besides, MPD (Maximum Percentile Distance) provided the shortest list of variables with the highest predictive accuracy.

Table 3 shows the confusion matrix rates (TPR, TNR, FPR, FNR) for all the binary classifications (comparisons 1 and 2), together with the false discovery rate (FDR), and the LOOCV predictive accuracy (ACC). No weights optimization has been performed in this case, that is, the weights correspond are the inverse of the prior variability of the prognostic variables (see Appendix 1 Supplementary Material). Table 4 shows the mean values of the three prognostic variables for the different groups of the confusion matrix, and the weights (W) used to define the distance criterion in the nearest-neighbor classifier.

Comparison 2.1: CR + PR vs. PD

In this comparison Serum Ferritin was the single selected prognostic variable (using MPD as feature selection) with a predictive accuracy of 95.82%. The SF mean value in the TN group (3288 ng/mL) is even higher than in the previous comparison, being the SF value in the TP group 266 ng/mL. Therefore, patients with Progressive Disease show a very clear inflammatory behavior as shown by the SF value at diagnosis. The TPR of this comparison is very high (98.43%), and the TNR is higher than in comparison 1 (22.22%). The TP and FP maximum and minimum SF values are closer to normal SF values (see table S1 of Supplementary Material). Conversely, the TN and FN corresponding SF signatures are extreme values.

Comparison 2.2: CR vs. PR

The best subset of prognostic variables for this case was found by MPD and was composed by SF, ALT and ALP, providing 92% of LOOCV predictive accuracy. The

TPR is very high (97.89%) and the TNR is quite low (11.76%), that is, the difference between partial and complete remission is very hard to tell, and the classifier tends to assign the complete remission class in most of the cases. There is a big gap between SF mean levels of both TP (249 ng/mL) and TN (2401 ng/mL) groups. Moreover, FP (405 ng/mL) and FN (2131 ng/mL) mean SF values are similar to the mean SF values of the TN and TP groups, respectively. The same happens with ALP, there is also a big difference between TP (116.8 U/L) and TN (376 U/L) values. The mean values in the FP (163.5 U/L) and FN (608.4 U/L) groups are also close to the TP and TN groups, which make those samples very difficult to hit using this k-NN classifier. SF and ALP have higher mean values in the TN group than in the TP group. However, in the case of ALT, the mean value in the TP group (23.7 U/L) is higher than in the TN group (18 U/L). Moreover, the difference between these two groups is very low. The ALT mean value in the FN group (74.4 U/L) is closer to the TP group, instead of being closer to the TN group, as it should be expected. This is due to the presence of some PR patients with anomalously large ALT values.

k-NN weights optimization

Particle Swarm Optimization (PSO) of the weights of the k-NN classifier was performed in order to improve TNR figures, giving more importance to the fact of hitting the negative ones, that is, increasing TNR while the overall accuracy is also improved (TPR is not affected). Details about PSO are given in Appendix 1 (see Supplementary Material).

Table 5 shows the TPR, TNR, FPR, FNR, FDR and Predictive Accuracy (Acc) obtained after weights optimization. TN rates were improved around 10% in comparisons 2.1, whilst in comparison 2.2 TP rate was improved around 1%. The overall accuracy was improved in all the cases around 1%. Table 6 shows the mean

values for TP, TN, FP, FN and the optimized weights for the prognostic variables (**W** weights). It can be observed that values of the weights increased after optimization for all the prognostic variables. Therefore, it is possible to improve the quality of the prediction, and minimize risk on the decisions, by optimizing the weights that are initially provided by the distance criterion.

CONCLUSIONS

In this paper we presented an optimum strategy to predict treatment response in HL. Three main discriminatory prognostic variables aroused in this analysis: Serum Ferritin, ALT and ALP.

Serum ferritin has been frequently used as a surrogate marker for systemic iron stores, but may be also elevated in specific circumstances without excess iron stores, such as in inflammation, correlating closely to the activity of the malignant lymphomas. Serum ferritin levels have been reported to be elevated in HL patients, in particular in advanced stages and during disease progression [16, 17]. Moreover, it has been proposed that the release of IL-6 stimulates the overproduction of hepcidin in the liver, which correlates with the iron restriction and contributes to anemia in HL [18]. In addition, the abundant microenvironment surrounding the neoplastic Hodgkin's and Reed-Sternberg cells may contribute to alterations in iron metabolism [19]. Besides, Serum Ferritin concentration closely follows the activity of the malignant lymphomas [20]. Other research work [21] has shown that levels of Serum Ferritin higher than 500 ng/mL is an important marker for predicting poor survival outcomes for non-Hodgkin Lymphoma. Nevertheless, and up to our knowledge, Serum Ferritin levels have not been yet related to the treatment response of HL patients.

Serum activity levels of ALT enzyme are routinely used as a biomarker of liver injury caused by drug toxicity, infection, alcohol and steatosis. ALT plays a key role in

the intermediary metabolism of glucose and amino acids, and also participates in cellular nitrogen metabolism and liver gluconeogenesis. This cytosolic enzyme catalyzes the transfer of the α-amino group from alanine to α-ketoglutaric acid. Serum levels of ALT are normally low (10 to 40 U/L), but any type of liver cell injury may modestly increase the ALT levels. Levels greater than 500 U/L occur most often in people with hepatic diseases, such as viral hepatitis, ischemic liver injury (shock liver), toxin-induced liver damage and tumor infiltration of liver. Despite the association between greatly elevated ALT levels and hepatocellular diseases, the levels of ALT does not correlate with the extent of liver cell damage [22].

The Alkaline Phosphatase test (ALP) is used to detect liver disease or bone disorders. In conditions affecting the liver, damaged liver cells release-increased amounts of ALP into the blood. Further, any condition that affects bone growth or causes increased activity of bone cells can affect ALP levels in the blood. In Non-Hodgkin Lymphomas, ALP is increased in patients with bone marrow affectation [23], thus reaching stage IV and worse prognosis. A recent study suggests that ALP together with Gamma-Glutamyl Transferase and Albumin may define advanced stages of HL [24]. Moreover, bone affectation is also associated to a high progression degree (HR: 1,96) [25]. However, in a patient with fever of unknown origin (FUO), highly elevated Alkaline Phosphatase and normal/slightly elevated serum transaminase levels suggest the possibility of lymphoma [26-28].

Overall, the results of this study shows that the combined use of these prognostic variables, SF, ALT and ALP, in a simple classifier allow to predict first line treatment response in HL patients with high accuracy; and confirms a close relationship between treatment response in HL, inflammation, iron overload and liver and bone damage. Particularly, the combination of feature selection methods (maximum percentile

distance), risk assessment analysis (ROC curve) and global optimization (PSO) provides biomarker discovery that is easily implemented in spreadsheet.

To conclude, detecting those HL patients who do not respond to the treatment at early stages may help improving their treatment. This study proposed a new prognostic analysis method, based on mathematical models that identify three simple prognostic variables currently gathered at diagnosis that may help detecting with high accuracy those HL patients with bad prognosis without any additional cost.

Acknowledgments

Enrique J. de Andrés was supported by the Spanish Ministerio de Economía y Competitividad (grant TIN2011-23558), and the medical analysis was supported by the Fondo de Investigaciones Sanitarias (Instituto Carlos III-grant PI12/01280). No other financial support has been received to perform this retrospective analysis.

REFERENCES

[1] Alvaro-Naranjo T, Lejeune M, Salvado-Usach MT, Bosch-Princep R, Reverter-Branchat G, Jaen-Martinez J, et al. Tumor-infiltrating cells as a prognostic factor in Hodgkin's lymphoma: a quantitative tissue microarray study in a large retrospective cohort of 267 patients. Leuk *Lymphoma* 2005; 46(11): 1581–91.

[2] Schreck S, Friebel D, Buettner M, Distel L, Grabenbauer G, Young LS, et al. Prognostic impact of tumour-infiltrating th2 and regulatory t cells in classical Hodgkin lymphoma. *Hematol Oncol* 2009; 27(1): 31–9.

[3] Hasenclever D, Diehl V, Armitage JO, Assouline D, Björkholm M, Brusamolino E, et al. A prognostic score for advanced Hodgkin's disease. *New Eng J Med* 1998; 339(21): 1506–14.

[4] Friedman S, Henry-Amar M, Cosset JM, Carde P, Hayat M, Dupouy N, et al. Evolution of erythrocyte sedimentation rate as predictor of early relapse in posttherapy early-stage Hodgkin's disease. *J Clin Oncol* 1988; 6(4): 596–602.

[5] Mauch P, Larson D, Osteen R, Silver B, Yeap B, Canellos G, et al. Prognostic factors for positive surgical staging in patients with Hodgkin's disease. *J Clin Oncol* 1990; 8(2): 257–65.

[6] Cheson BD. New staging and response criteria for non-Hodgkin lymphoma and Hodgkin lymphoma. *Radiol Clin North Am* 2008; 46(2): 213–23.

[7] Biggi A, Gallamini A, Chauvie S, Hutchings M, Kostakoglu L, Gregianin M, et al. International validation study for interim pet in abvd-treated, advanced-stage Hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. *J Nucl Med* 2013; 54(5): 683–90.

[8] Smolewski P, Robak T, Krykowski E, Blasinska-Morawiec M, Niewiadomska H, Pluzanska A, et al. Prognostic factors in Hodgkin's disease: multivariate analysis of 327 patients from a single institution. *Clin Cancer Res* 2000; 6(3): 1150–60.

[9] Zander T, Wiedenmann S and Wolf J. Prognostic factors in Hodgkin's lymphoma. *Ann Oncol* 2002; 13 Suppl 1:67–74.

[10] Josting A. Prognostic factors in Hodgkin lymphoma. *Expert Rev Hematol* 2010; 3(5): 583–92.

[11] Provencio M, Espana P, Millan I, Yebra M, Sanchez AC, de la Torre A, et al. Prognostic factors in Hodgkin's disease. *Leuk Lymphoma* 2004; 45(6): 1133– 9.

[12] Ong DST, Wang L, Zhu Y, Ho B and Ding JL. The response of ferritin to lps and acute phase of pseudomonas infection. *J Endotoxin Res* 2005; 11(5): 267–80.

[13] Cheson BD. The international harmonization project for response criteria in lymphoma clinical trials. *Hematol Oncol Clin North Am* 2007; 21(5): 841–54.

[14] Saligan L, Fernández-Martínez JL, deAndrés-Galiana EJ and Sonis S. Supervised classification by filter methods and recursive feature elimination predicts risk of radiotherapy-related fatigue in patients with prostate cancer. *Cancer Informatics*, 2014; 13:141–152, 12.

[15] Gandhi MK, Lambley E, Burrows J, Dua U, Elliott S, Shaw PJ, et al. Plasma epstein-barr virus (EBV) DNA is a biomarker for EBV-positive Hodgkin's lymphoma. *Clin Cancer Res* 2006; 12(2): 460–4.

[16] Bezwoda WR, Derman DP, Bothwell TH, Baynes R, Hesdorffer C and MacPhail AP. Serum ferritin and Hodgkin's disease. *Scand J Haematol* 1985; 35:505–510.

[17] Dörner MH, Abel U, Fritze D, Manke HG and Drings P. Serum ferritin in relation to the course of Hodgkin's disease. *Cancer* 1983; 52:2308–2312

[18] Hohaus S, Massini G, Giachelia M, Vannata B, Bozzoli V, Cuccaro A, et al. Anemia in Hodgkin's lymphoma: the role of interleukin-6 and hepcidin. *J Clin Oncol* 2010; 28(15): 2538–43.

[19] Hohaus S, Giachelia M, Cuccaro A, Voso MT and Leone G. Iron in Hodgkin's lymphoma. *Crit Rev Oncog* 2013; 18(5): 463–9.

[20] Aulbert E and Steffens O. Serum ferritin–a tumor marker in malignant lymphomas? *Onkologie* 1990; 13(2): 102–8.

[21] Yoh KA, Lee HS, Park LC, Lee EM, Shin SH, Park DJ, et al. The prognostic significance of elevated levels of serum ferritin before chemotherapy in patients with non-Hodgkin lymphoma. *Clin Lymphoma Myeloma Leuk* 2014; 14(1): 43–9.

[22] Kaplan MM. Alanine aminotransferase levels: what's normal? *Ann Intern Med* 2002; 137(1): 49–51.

[23] Kittivorapart J and Chinthammitr Y. Incidence and risk factors of bone marrow involvement by non-Hodgkin lymphoma. *J Med Assoc Thai* 2011; 94 Suppl 1: S239–45.

[24] Jamakovic M and Baljic R. Significance of copper level in serum and routine laboratory parameters in estimation of outspreading of Hodgkin's lymphoma. *Med Arch* 2013; 67(3): 185–7.

[25] El-Galaly TC, Hutchings M, Mylam KJ, Brown Pde N, Bukh A, Johnsen HE, et al. Impact of f-fluorodeoxyglucose positron emission tomography/computed tomography staging in newly diagnosed classical Hodgkin lymphoma: fewer cases with stage I disease and more with skeletal involvement. *Leuk Lymphoma* 2014.

[26] Brensilver HL and Kaplan MM. Significance of elevated liver alkaline phosphatase in serum. *Gastroenterology* 1975; 68(6): 1556–62.

[27] Brinckmeyer LM, Skovsgaard T, Thiede T, Vesterager L and Nissen NI. The liver in Hodgkin's disease. Clinico-pathological relations. *Eur J Cancer Clin Oncol* 1982; 18(5): 421–8.

[28] Cunha BA. Fever of unknown origin (FUO): diagnostic importance of serum ferritin levels. *Scand J Infect Dis* 2007; 39(6-7): 651–2.

LIST OF CAPTIONS

Figure 1: Flow diagram for the Hodgkin Lymphoma treatment response prediction model. The methodology is composed of 3 steps: 1. Filtering and Imputing data. 2. Feature Selection. 3. k-NN design and risk analysis. In each box the different sub steps are also detailed.