Multiple Myeloma data description

Rima GUILAL[1,*], Ahmed Fouad BENDAHMANE[2], Nesma SETTOUTI[1], Mohammed El Amine CHIKH [1], Naima MESLI[2]

[1] Laboratory of Biomedical Engineering, Faculty of Technology, University of Tlemcen, Algeria
[2] Hematology department, University hospital of TLEMCEN, Algeria
* rima.guilal@univ-tlemcen.dz

1 Domain Background

The Multiple Myeloma (MM) is a type of blood cancer that affects the plasma cells in bone morrow [1], it presents 2% of all cancers and 10% of hematologic cancers. The MM is one of several complicated cancers, often, no symptoms are detected in the early stage [2]. Once tumor has developed, the suspicion of MM requires the observation of following symptoms which are described with acronym "CRAB":

- High **C**alcium levels in the blood (Hypercalcemia).
- **R**enal or Kidney problems (Kidney failure).
- **A**nemia.
- **B**one pain and Osteolytic lesions.

1.1 Multiple Myeloma diagnosis

Its diagnosis based on different medical exams and tests including:

- **Hematology tests**: are all medical tests concerning the study of blood-forming organs involved in bleeding and coagulation. It includes: CBC (complete blood count), peripheral blood smear.
- **Cytology exams**: are useful for diagnosis of MM in order to detect the abnormal blood cells. Myelogram exam is one of those medical exams which is used to quantify the plasma cells rate ($> 10\%$), to identify and detect those that are abnormal. In some cases, the bone marrow biopsy is often necessary in a second step, to confirm the diagnosis.

- **Protein tests**: allow to measure the total amount of protein in the blood or urea. It the first step in diagnosing of Multiple Myeloma. Serum protein electrophoresis is used to measure the blood levels of different antibodies. Another test called protein immunofixation should be used to determine the type of abnormal antibody (IgG, IgA, IgG, IgE, IgM). In some cases, fragments of abnormal myeloma protein are filtered by the kidney into urine, so the urine protein tests are used to find monoclonal antibody in urine (Bence-Jones protein, urine protein electrophoresis and urine immunofixation).

- **Medical imaging tests**: the myeloma cells produce a substance that causes bone destruction. For that, it's necessary to use imaging tests in order to detect lytic bone lesions. The most used test is standard X-ray.

- **Blood biology and chemistry tests**: are carried out in order to evaluate the function organs and also to detect abnormalities (calcemia, kidney function tests, blood ionogram, serology, liver function tests, glucose level, hemostatic mechanisms, ... etc.).

1.2 Multiple Myeloma clinical forms

There are different forms of Multiple Myeloma according to the diagnostic criteria:

- **Monoclonal gammopathy of undetermined significance** (MGUS) [3] is a precancerous condition in which an abnormal protein is in blood. Usually, it causes no problems, but sometimes it can progress over years to other disorders, including some forms of blood cancer (lympho proliferative disorder, Multiple Myeloma, ...).

- **Smoldering Multiple Myeloma** is an early form of MM, it is also called **asymptomatic myeloma**, which means the absence of clinical and biological criteria of symptomatic Multiple Myeloma. it is defined by the presence of a monoclonal protein in the serum or plasmocytic proliferation. This type of myeloma is between MGUS and active Multiple Myeloma (symptomatic).

- **Active Multiple Myeloma** also called **symptomatic myeloma**, people with active Multiple Myeloma have symptoms. It is progressive and needs to be treated.

2 Problem Statement

It is necessary to determinate the stage of the disease. There are three staging systems for Multiple Myeloma that differ in the evaluation factors:

A - **Durie-Selmon classification:** used since 1975 [4], it divides MM into three stages indicated by the Roman numerals (stage I, stage II, stage III). Each stage is also classified into A or B, depending on kidney function (see Table 1).

Stages	**Criteria**
Stage I	One or more of the following criteria: - Hg (Hemoglobin) $>$ 10 g/dl. - Normal calcium level. - Bone X-ray is normal or only one lytic bone lesion. - Low level of monoclonal proteins.
Stage II	Neither stage I nor stage III
Stage III	One or more of the following criteria: - Hg $<$ 8,5 g/dl - calcium level $>$ 12 mg/dl - lytic bone lesions (more than one) - High level of monoclonal proteins.
Sub-classification	
A	creatinine level $<$20 mg/l (no kidney problems)
B	creatinine level $>$20 mg/l (Renal or Kidney problems (Kidney failure))

Table 1: Durie-Selmon classification

B - **International Staging System (ISS):** used commonly since 2005, [5] is based on two factors: the measure of serum Albumin and the levels of serum Beta-2-microglobulin (see Table 2).

Stages	**Criteria**
Stage I	B2M $<$3.5 mg/l alb $\geq$ 35 g/l
Stage II	B2M $<$3.5 mg/l and alb $\geq$ 35 g/l Or B2M 3.5 – 5.5 mg/l
Stage III	B2M $>$ 5.5 mg/l

Table 2: International Staging System (ISS)

C - **Revised International Staging System (R-ISS)** is a new risk-stratification algorithm with improved prognostic power compared to the previously staging methods [6]. It incorporates the following steps:

- 1^{st} STEP: Determine the ISS stage of the patient (see Table 2).
- 2^{nd} STEP: Assess risk according to chromosomal abnormalities using FISH [1] (see Table 3)

[1] Fluorescence in situ hybridization (FISH) is a molecular cytogenetic technique that uses

Risk	Criteria
Standard Risk	No high-risk chromosomal abnormalities
High Risk	Presence of del(17p), and/or translocation t(4;14), and/or translocation t(14,16)

Table 3: Risk according to chromosomal abnormalities by FISH

– 3^{rd} STEP: Assess risk according to serum lactate dehydrogenase (LDH) level (see Table 4)

Risk	Criteria
Normal	<to the normal defined by the laboratory
High	>to the normal defined by the laboratory

Table 4: Risk according to LDH level

– 4^{th} STEP: determine the R-ISS stage according to the three steps above (see Table 5)

R-ISS Stage	Criteria
Stage I	ISS stage I and standard-risk chromosomal abnormalities and normal LDH
Stage II	Not R-ISS stage I or III
Stage III	ISS stage III and high-risk chromosomal abnormalities or high LDH

Table 5: Criteria of the revised international staging system

3 Description of the dataset

The MM dataset was collected at the cancer center of University hospital of TLEMCEN, Algeria [2]. It consists of 200 patients who are diagnosed during the period 2008-2019, and 57 features including cover demographic information, personnel and family antecedents, different results of medical exams and tests diagnosis of MM.

to detect and localize the presence or absence of specific DNA sequences on chromosomes.

[2] http://www.chu-tlemcen.dz/

The output class contained the labels stage of MM cancer which are classified by specialists on Hematology using Durie-Selmon staging and International Staging System. The features information and their code in the dataset are detailed in Table 6 below.

Table 6: DataSet description

Information	Code	Attribute	Type	Min	Max
Demographic information	gender	Gender	Nominal		
	age	Age	Numeric	38	98
	city	Wilaya of residence	Nominal		
	married	Married	Boolean		
	nbr_child	number of children	Numeric	0	12
Clinical Exam	weight	Weight	Numeric	40	96
	body_surf	Body surface	Numeric	1.09	2.12
	blood	Blood type	Nominal		
	asth&bone	Asthenia and bone pain	Boolean		
	anemia	clinical signs of anemia	Boolean		
Personal and Family Antecedents	HBP	Antecedent_HBP	Boolean		
	diabete	Antecedent_diabete	Boolean		
	tobbacco	Antecedent_tobbacco	Boolean		
	chron_disea	Chronic family diseases	Boolean		
	hrd_blo_disea	Hereditary blood diseases	Boolean		
Hematology Tests	CBC_WBC	rate of white blood cells	Numeric	1.7	23.5
	CBC_RBC	rate of red blood cells	Numeric	1.1	5.88
	CBC_plats	rate of platelets	Numeric	66	692
	CBC_Hgb	Hemoglobin level	Numeric	3.8	17.1
	CBC_Hct	Hematocrit	Numeric	11.3	52.61
	CBC_MCV	Mean corpuscular volume	Numeric	27	120
	CBC_MCHC	Mean corpuscular hemoglobin concentration	Numeric	27.5	39.2
Cytology Exams	roll_RBC	red blood cells go on a roll	Boolean		
	plasma_cells	the rate of plasma cells	Numeric	0	95
Proteins Tests	B2M	β-2 Microglobulin test	Numeric	1.36	31.95
	prot_rate	proteins rate	Numeric	48	168
	alb	Albumin	Numeric	13.52	49.3
	α_glob	α_globulin	Numeric	2.55	36.92
	β_glob	β_globulin	Numeric	2.7	93.3
	γ_glob	γ_globulin	Numeric	0	119
	BJp	Bence Jonce protein	Boolean		
	24h_prot	24 hour proteinuria	Boolean		
	Ig	Abnormal Imunoglobulin	Nominal		
	chain	Type of free light chain	Nominal		

Continued on next page

Medical imaging	ost_les	osteolytic lesions	Boolean		
Blood (and urine) Biology and chemistry Tests	VS	Erythrocyte sedimentation rate (ESR)	Numeric	2	170
	Ca	Calcium test	Numeric	10.3	211.74
	K	Potassium test	Numeric	2.5	9.79
	Na	Sodium test	Numeric	120	154.2
	P	Phosphorus test	Numeric	1.68	114
	CRP	C-reactive protein	Boolean		
	creat	Creatinine level	Numeric	1.49	178
	urea	Urea	Numeric	0.12	13.9
	clair_creat	creatinine clearance test	Numeric	0.49	171.92
	SGOT	Serum Glutamic Oxaloacetic Transaminase	Numeric	1	67
	SGPT	Serum Glutamic-Pyruvic Transaminase	Numeric	1.28	125
	GGT	Gamma-glutamylTransferase	Numeric	4.3	586
	PAL	Alkaline phosphatase (ALK)	Numeric	2.34	3348
	Ac_Anti_HCV	Anti-hepatitis C antibodies test	Boolean		
	HIV	Human immunodeficiency virus test	Boolean		
	Ag_HBS	Hepatitis-B Surface Antigen test	Boolean		
	gly	Blood glucose test	Numeric	0.4	4.6
	TCA	Activated thromboplastin time (ATT)	Numeric	3	80.2
	TP	Prothrombin rate	Numeric	25.5	100
	Fib	Fibrinogen blood test	Numeric	1.64	200
	Ferr	Ferritin rate in blood	Numeric	6.36	1176
	LDH	Lactate dehydrogenase	Numeric	68	728
	cardio_EF	Left Ventricular Ejection Fraction (LVEF)	Numeric	44.43	88

As well as, the Table 7 presents output classes with the number of examples for each class.

Code	Class	# Exemples
MGUS	Monoclonal gammopathy of undetermined significance	5
ASYM	Asymptomatic Myeloma	5
IA	stage I type A	12
IB	stage I type B	1
IIA	stage II type A	8
IIB	stage II type B	1
IIIA	stage III type A	98
IIIB	stage III type B	32
PLASMO	Plasmocytome	2

Table 7: Description of MM classes

4 Solution Statement

In literature, all recent works which have been proposed to the assistance with the medical diagnosis in multiple myeloma (MM) disease, are based on genetic databases [7], [8], [9], [10]. For this reason, we collect a new dataset that contains results of different diagnosis exams of MM.

This new dataset can be used to display the Multiple Myeloma diagnosis problem as a clinical and para-clinical factors detection by using features selection methods.

In other words, we can also use it to solve imbalanced data classification problems based on supervised machine learning with multi-classes output.

As a future works, we will try to enrich our dataset with healthy subjects.

References

[1] L. Furchtgott, A. Bolomsky, F. Gruber, M. K. Samur, J. J. Keats, J. Yesil, K. Stangelberger, M. Attal, P. Moreau, H. Avet-Loiseau, K. Runge, D. Wuest, K. Rich, I. Khalil, B. Hayete, H. Ludwig, N. Munshi, and D. Auclair, "Multiple myeloma drivers of high risk and response to stem cell transplantation identified by causal machine learning: Out-of-cohort and experimental validation," *Blood*, vol. 130, no. Suppl 1, pp. 3029–3029, 2017. [Online]. Available: http://www.bloodjournal.org/content/130/Suppl_1/3029

[2] "American cancer society," https://www.cancer.org/cancer/multiple-myeloma/detection-diagnosis-staging/detection, [Consulted on line 14-May-2019]. [Online]. Available: https://www.cancer.org/cancer/multiple-myeloma/detection-diagnosis-staging/detection

[3] T. H. Mouhieddine, L. D. Weeks, and I. M. Ghobrial, "Monoclonal gammopathy of undetermined significance (mgus)," *Blood*, pp. blood–2 019 846 782, 2019.

[4] G. Filonzi, K. Mancuso, E. Zamagni, C. Nanni, P. Spinnato, M. Cavo, S. Fanti, E. Salizzoni, and A. Bazzocchi, "A comparison of different staging systems for multiple myeloma: can the mri pattern play a prognostic role?" *American Journal of Roentgenology*, vol. 209, no. 1, pp. 152–158, 2017.

[5] P. R. Greipp, J. S. Miguel, B. G. Durie, J. J. Crowley, B. Barlogie, J. Bladé, M. Boccadoro, J. A. Child, H. Avet-Loiseau, R. A. Kyle *et al.*, "International staging system for multiple myeloma," *Journal of clinical oncology*, vol. 23, no. 15, pp. 3412–3420, 2005.

[6] A. Palumbo, H. Avet-Loiseau, S. Oliva, H. M. Lokhorst, H. Goldschmidt, L. Rosinol, P. Richardson, S. Caltagirone, J. J. Lahuerta, T. Facon *et al.*, "Revised international staging system for multiple myeloma: a report from international myeloma working group," *Journal of clinical oncology*, vol. 33, no. 26, p. 2863, 2015.

[7] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8–17, 2015.

[8] A. Palmaro, M. Gauthier, C. Conte, P. Grosclaude, F. Despas, and M. Lapeyre-Mestre, "Identifying multiple myeloma patients using data from the french health insurance databases: Validation using a cancer registry," *Medicine*, vol. 96, no. 12, p. 8, 2017.

[9] J. Ubels, P. Sonneveld, E. H. van Beers, A. Broijl, M. H. van Vliet, and J. de Ridder, "Predicting treatment benefit in multiple myeloma through simulation of alternative treatment effects," *Nature communications*, vol. 9, p. 10, 2018.

[10] M. Rossi, M. Martino, P. Guzzi, P. Tagliaferri, and P. Tassone, "New approaches to predict outcome and personalize therapy in multiple myeloma: from micrornas to integrated genomics," *Ann Hematol Oncol*, vol. 2, p. 5, 2015.