Discovery of Characteristics of Patients with Increased Level of Inflammation

Ljiljana Trtica Majnarić[1]*, Pinar Yildirim[2] and Andreas Holzinger[3]

[1]*Department of Family Medicine, School of Medicine, University of Osijek, Croatia*
[2]*Department of Computer Engineering, Faculty of Engineering and Architecture, Okan University, Istanbul, Turkey*
[3]*Institute of Information Systems and Computer Media, Graz University of Technology, Graz, Austria*

Abstract

This paper is a study on knowledge discovery for the prediction of characteristics of older patients with increased level of inflammation. The etiology of inflammation is thought to be multifactorial and associated with the development of chronic aging diseases. Chronic low grade inflammation, expressed by slightly elevated serum concentrations of the inflammatory marker C-reactive protein (CRP), has been showed to be associated with increased frailty and overall and specific cardiovascular and noncardiovascular mortality. However, it has not been discovered before which conditions, taken all at once, are associated with elevated systems level of inflammation. To answer this question, we used the group of older primary health care attenders, burdened with multiple chronic conditions and described their health status by many aspects. The dataset was composed of 61 low-cost health parameters, many of which were data from patient health records. To predict the characteristics of patients with increased level of inflammation, we used a Linear Regression model and compared the results with some classfication algorithms. In this way, we selected 11 relevant predictors of inflammation and explained their meaning according to the existing knowledge. We could realise that many of them represent the components of the highly conserved functional network in which inflammation is the intermediate mechanism, linking these components together. These components include the metabolic, the neuroendocrine and the immune system, proposed as influencing each other during the development of age-related chronic diseases. We have also identified some new components, represented by the parameters indicating inflammation-mediated locomotor system disorders and the pituitary hormone prolactin serum concentration variations. This model, resulted from the knowledge discovery procedure, can be used to provide guidelines for further research on chronic low grade inflammation and for more practical purposes, to help physicians recognizing older persons who are at increased risk for frailty and death.

Introduction

Inflammation is a well conserved set of responses by the tissues and the body to tissue injury or infection, aimed at returning physiological homeostasis and tissue repair. However, when uncontrolled or chronically maintained, it can also be detrimental and contribute to disease development [1]. During inflammation, the inflammatory cytokines: tumor necrosis factor-α (TNF-α), interleukin-1 (IL-1) and interleukin-6 (IL-6), are secreted to enhance inflammation and help control its course. In many inflammation-related activities, these cytokines show redundancy in action. Specifically, stimulation of acute-phase proteins production in the liver and activation of the hypothalamic-pituitary-adrenal stress axis, are two functions which are almost completely controlled by secretion of the cytokine IL-6. This cytokine is also interesting for its marked pleiotropy and involvement in the regulation of a variety of physiologic processes, including hematologic, immune, neurologic, endocrine and metabolic functions [2].

Chronic low grade inflammation, expressed by slightly elevated serum concentrations of inflammatory markers, interleukin-6 (IL-6) and C-reactive protein (CRP), both known as acute phase-reactants, has for a long time been recognised as an independent cardiovascular (CV) risk factor. Especially CRP has been characterised with good prognostic properties [3]. These findings have been showed constant across various populations, including initially healthy individuals, those with marked classical CV risk factors, elderly people and those with clinically manifest CV disease [4,5]. In this regard, positive associations have been found between CRP and body mass index (BMI), fasting glucose and insulin serum concentrations, systolic blood pressure and serum lipids, indicating the involvement of inflammation

in the pathogenesis of the insulin resistance (the metabolic) syndrome [6]. This is the complex trait which involves obesity of abdominal (visceral) type and is known to substantially increase the susceptibility for both, diabetes and CV disease. Prevalence of this syndrome increases with age [7]. More generally (Figure 1), evidence suggests that during aging, the tight regulation of inflammatory genes becomes less effective. As the consequence, soluble markers of inflammation, notably CRP, are measurable even when there is no apparent evidence of infection, trauma, or other stressful conditions [8,9]. The etiology is thought to be multifactorial. Some evidence indicates the particular importance of the cumulative antigenic stimulation, along with the loss of immunoregulatory functions in older age. Some other evidence indicates age-related decline in sex hormones, estrogen in women and testosterone in men [8,10]. Recent studies also highlight the importance of the genetic variation in IL-6 gene expression, for variation in serum concentrations of IL-6 and CRP, observed among individuals in the population [11]. The state of chronic low grade inflammation has been considered responsible for the initiation/progression of the main aging chronic diseases, including CV disease, Alzheimer`s and other dementia, osteoporosis, autoimmune and lymphoproliferative diseases [2,8,10]. In addition, elevated systems levels of inflammation have been

*Corresponding author: Ljiljana Trtica Majnarić, Assistant Professor, Head of the Family Medicine Department, School of Medicine, University of Osijek, Croatia, Tel: +385916136133; E-mail: ljiljana.majnaric@gmail.com

Figure 1: Chronic low grade inflammation.
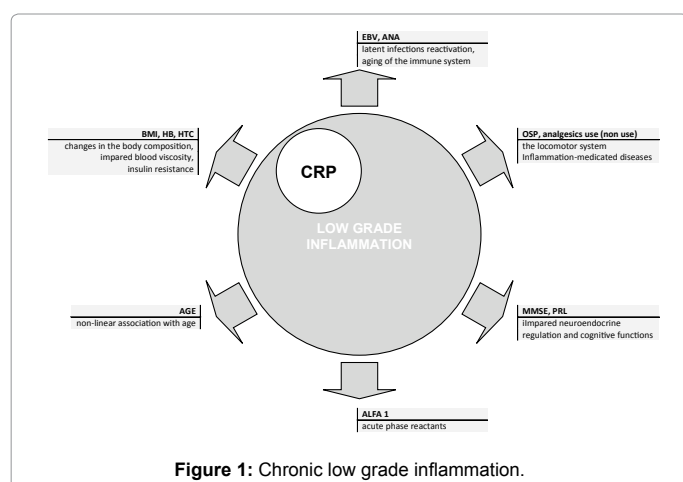
proposed to play a major role in the development, in older people, of the frailty syndrome, found to be associated with their increased vulnerability for disease and death. The syndrome is characterised with certain phenotypes typically seen in people of advanced age, including decreased lean body mass, osteopenia, anemia and low cholesterol serum concentrations [10]. Finally, in more general means, associations have been found between elevated IL-6 and CRP serum concentrations and increased overall and specific cardiovascular and noncardiovascular mortality rates. However, it has not been clearly stated, yet, whether these inflammation markers are merely reflective of established chronic diseases, or whether they also stimulate their development [5].

In the group of older, community-dwelling patients, burdened with chronic medical conditions, we wanted to find out hidden relationships in their health data and to link them to elevated serum concentrations of CRP, a well confirmed marker of inflammation. For this purpose, we used a dataset in which the health status of subjects was determined systematically, by many aspects. The model resulted from the knowledge discovery procedure and the selected predictors can be used to provide guidelines for further research and for more practical purposes, to help physicians to recognize older persons, primary health care attenders, who are at increased risk for frailty and death.

Related Work

There were several studies performed for knowledge discovery on C-reactive protein (CRP), but no one comprehensively designed. The first example. Dahlstrom et al. examined whether representation of physicians' expert knowledge in a simple heuristic model can improve data mining methods in prognostic assessment of patients with rheumatoid arthritis (RA). They interviewed consultants and considered several prognostic indicators which they suggested. Then they carried out a clustering analysis by using a k-means algorithm and some of these suggested variables, such as the erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), the number of swollen joints and the number of tender joints. They identified some prognostic subgroups and found that there is a positive correlation between ESR and CRP in this patient setting [12]. The second example. CRP, a marker of inflammation, has been identified as a risk factor for cardiovascular disease and mortality. Alley et al. used data on adults aged 20 and more from the fourth National Health and Nutrition Examination Survey, a nationally representative cross-sectional survey, and applied statistical methods, T tests and multinomial logistic regression, to examine the association between socioeconomic status and CRP in US adults. They found these associations to occur only at high levels of CRP, while there was no significant difference in the prevalence of moderate or high

values of CRP according to the differences in socioeconomic status [13].

Materials and Methods

The sample

The study was conducted in a family practice in an urban area (the town of Osijek, the north-eastern part of Croatia). The region is known for high prevalence of cardiovascular and other chronic disorders, above the average of Croatia [14]. The sample consisted of 93 subjects, aged 50 years and more, who gave their consent. There were 35 males and 58 females, 50-89 years old (median 69), burdened with chronic medical conditions. The study protocol was approved by the local ethics committee.

Parameters description and the modelling process

For the purpose of this study, health data were collected systematically, to determine many aspects of the health-status of patients. Only not costly, easy available parameters were used, a large proportion of which were routinely collected parameters used from patients' health records. There was a total number of 61 parameters (Tables 1-3). Nominal parameters indicated age and sex, diagnoses of the main chronic diseases, information on drugs use and anthropometric measures. A substantial proportion of subjects were burdened with different chronic disorders (Table 4). A large number of haematological and biochemical laboratory tests were performed, to determine the age-related pathophysiologic changes. These tests included parameters indicating: 1) inflammation (parameters: LE, NEU, EO, MO, LY, CRP, ALFA1 and 2, BETA, GAMA), 2) the nutritional status (parameters: E, HB, HTC, MCV, FE, PROT, ALB, VITB12, FOLNA, HOMCIS), 3) the metabolic status (FGlu, HbA1c, Chol, TG, HDL, INS), 4) chronic renal impairment (parameters: Clear), 5) latent chronic infections (parameters: CMV, EBV, HBG, HPA), 6) impaired humoral immunity (parameters: RF, IGE, ANA) and 7) the neuroendocrine status (parameters: CORTIS, TSH, FT3, FT4, PRL) (Tables 2 and 3).

No	Attribute	Description	Type
1	Age	Age (years)	Numeric
2	Sex	(M=Male, F=Female)	Nominal
3	Hyper	Hypertension (yes, no)	Nominal
4	DM	Diabetes mellitus (yes, IGT=Impaired glucose tolerance, No)	Nominal
5	F Glu	Fasting blood glucose (mmol/L)	Numeric
6	HbA1c	Glycosilated Haemoglobin (%) (showing average blood glucose during last three months)	Numeric
7	Chol	Total Cholesterol (mmol/L)	Numeric
8	TG	Triglycerides (mmol/L)	Numeric
9	HDL	HDL-cholesterol (mmol/L)	Numeric
10	Statins	Therapy with statins (yes, no)	Nominal
11	Anticoag	Therapy with anticoagulant/antiaggregant drugs (yes, no)	Nominal
12	CVD	Cardiovascular diseases (yes, no) (myocardial infarction, angina, history of revascularisation, stroke, transient ischaemic cerebral event, peripheral vascular disease)	Nominal
13	BMI	Body Mass Index (kg/m²)	Numeric
14	W/h	Waist/hip ratio	Numeric
15	Arm cir	Mid arm circumference (mm)	Nominal
16	Skinf	Triceps skinfold thickness (mm)	integer
17	Gastro	Gastroduodenal disorders (yes, no) (gastritis, ulcer)	Nominal
18	Uro	Chronic urinary tract disorders (yes, no) (recurrent cystitis in women, symptoms of prostatism in men)	Nominal
19	COPB	Chronic obstructive pulmonary disease (yes, no)	Nominal
20	Aller d	Allergy (Rhinitis and/or Asthma) (yes, no)	Nominal

Table 1: The attributes used in the dataset (1-20).

No	Attribute	Description	Type
21	Dr aller	Drugs allergy (yes, no)	Nominal
22	Analg	Therapy with analgetics/non-steroidal anti-inflammatory drugs (yes, no) Patients with severe osteo-arthritis	Nominal
23	Derm	Chronic skin disorders (yes, no) (chronic dermatitis, dermatomycosis)	Nominal
24	Neo	Malignancy, including skin malignancy (yes, no)	Nominal
25	OSP	Osteoporosis (yes, no)	Nominal
26	Psy	Neuropsychiatric disorders (yes, no) (anxiety/depression, Parkinson`s disease, cognitive impairments)	Nominal
27	MMS	Mini Mental Score – test for screening on cognitive dysfunction Max Score=30 Score <24 indicates cognitive impairment	integer
28	CMV	Cytomegalovirus specific IgG antibodies (IU/ml)	Numeric
29	EBV	Epstein-Barr virus specific IgG (IU/ml)	Numeric
30	HBG	Helicobacter pylori specific IgG (IU/ml)	Numeric
31	HPA	Helicobacter pylori specific IgA (IU/ml)	Numeric
32	LE	Leukocytes Number $\times 10^9$/L	Numeric
33	NEU	Neutrophils % in White Blood Cell differential	Numeric
34	EO	Eosinophils % in White Blood Cell differential	Numeric
35	MO	Monocytes % in White Blood Cell differential	Numeric
36	LY	Lymphocytes % in White Blood Cell differential	Numeric
37	E	Erythrocytes number $\times 10^{12}$/L	Numeric
38	HB	Haemoglobin (g/L)	integer
39	HTC	Haematocrite (erythrocyte volume blood fraction)	Numeric
40	MCV	Mean cell Volume (fL)	Numeric

Table 2: The attributes used in the dataset (21-40).

No	Attribute	Description	Type
41	FE	Iron (g/L)	Numeric
42	PROT	Total serum proteins (g/L)	Nominal
43	ALB	Serum albumin (g/L)	Numeric
44	Clear	Creatinine clearance (ml/s/1.73 m²)	Nominal
45	HOMCIS	Homocistein (μmol/L)	Numeric
46	ALFA1	Serum protein electrophoresis (g/L)	Numeric
47	ALFA2	Serum protein electrophoresis (g/L)	Numeric
48	BETA	Serum protein electrophoresis (g/L)	Numeric
49	GAMA	Serum protein electrophoresis (g/L)	Numeric
50	RF	Rheumatoid Factor level (IU/ml)	Numeric
51	VITB12	Vitamin B12 (pmol/L)	Numeric
52	FOLNA	Folic acid (mM/L)	Numeric
53	INS	Insulin (μIU/L)	Numeric
54	CORTIS	Cortisol in the morning (nmol/L)	Numeric
55	PRL	Prolactin in the morning (mIU/L)	Numeric
56	TSH	Thyroid-stimulating hormone (IU/ml)	Nominal
57	FT3	Free triiodothyronine (pmol/L)	Numeric
58	FT4	Free thyroxine (pmol/L)	Numeric
59	ANA	Antinuclear antibodies (autoantibodies) (μIU/ml)	Numeric
60	IGE	IgE (kIU/L)	Numeric
61	CRP	C-reactive protein (mg/L)	Nominal

Table 3: The attributes used in the dataset (41-61).

The same dataset, although with minimal corrections in variables selection, we have already used for solving other complex research tasks [15-17]. We wanted to show that by collecting health data systematically and by using data mining techniques for their analysis, it is possible to elucidate important clinical contexts of some complex age-related medical conditions, usually presented with multi-morbidity (the simultaneous presence of several disorders at the same individual).

The modeling procedure can reveal many important elements of the networked pathophysiology reaction. The results can be used for prediction purposes, to support medical doctors in their decision making, or to accelerate research, by mapping the main components and relationships within the network. In this terms, this procedure may be recommended as a new integrative approach, with the potential to add value to much more complex and time and cost consuming systems methods [18].

Linear Regression

Regression analysis is an important statistical method for the analysis of medical data. It enables the identification and characterization of relationships among multiple factors [19]. Linear regression is a well known method of mathematical modeling of the relationship between a dependent variable and one or more independent variables. Regression uses existing (or known) values to forecast the required parameters [20]. The goal of this model, is to predict the response to n data points $(x_1, y_1),(x_2,y_2),…,(x_n,y_n)$ by a regression model given by

$$y=a_0+a_1x$$

Where a_0 and a_1 are the constants of the regression model. A measure of goodness of fit, that is, how well a_0+a_1x predicts the response variable y is the magnitude of the residual ε_i at each of the n data points.

$$\varepsilon_i=y_i-(a_0+a_1x_i)$$

Ideally, if all the residuals ε_i are zero, one may find an equation in which all the points lie on the model. Thus, minimization of the residual is an objective of obtaining regression coefficients.

The most popular method to minimize the residual is the least squares methods, where the estimates of the constants of the models are chosen such that the sum of the squared residuals is minimized, that is minimize $\sum_{i=1}^{n}\varepsilon_i^2$ [20]. Measures of association provide an initial impression of the extent of statistical dependence between variables. If the dependent and independent variables are continuous, a correlation coefficient (r) can be calculated as a measure of the strength of the relationship between them.

In many cases, the contribution of a single independent variable does not alone suffice to explain the dependent variable Y. If this is so, one can perform a multivariate linear regression to study the effect of multiple variables on the dependent variable [19].

In the multivariate regression model, the dependent variable is described as a linear function of the independent variables x_i, as follows:

$$y=a+b_1x_1+b_2x_2+\ldots+b_nx_n$$

The model permits the computation of a regression coefficient b_i for each independent variable. Just as in univariable regression, the coefficient of determination describes the overall relationship between the independent variables x_i and the dependent variable y. It corresponds to the square of the multiple correlation coefficients, which is the correlation between y and $b_1x_1+b_2x_2+\ldots+b_nx_n$ [19].

Decision Table

Decision Table summarizes the dataset with a 'decision table', a decision table contains the same number of attributes as the original dataset, and a new data item is assigned a category by finding the line in the decision table that matches the non-class values of the data item. This implementation employs the wrapper method to find a good subset of attributes for inclusion in the table. By eliminating attributes that contribute little or nothing to a model of the dataset, the algorithm reduces the likelihood of over-fitting and creates a smaller, more condensed decision table [21-23].

Kstar

K is an instance-based classifier, the class of a test instance is based upon the class of those training instances similar to it, as determined by some similarity function. The underlying assumption of instance-based classifiers is such as K [23,24].

M5Rules

The method for generating rules from model trees, which is called M5rules, is straightforward and works as follows: a tree learner is applied to the full training dataset and a pruned tree is learned. Next, the best branch (according to some heuristic) is made into a rule and the tree is discarded. All instances covered by the rule are removed from the dataset. The process is applied recursively to the remaining instances and terminates when all instances are covered by one or more rules. This is a basic separate-and-conquer strategy for learning rules: however, instead of building a single rule, as it is usually done, we built a full model tree at each stage, and make its "best" branch into a rule. This avoids potential for over-pruning called hasty generalization. In contrast to PART (Partial Decision Trees) that employs the same strategy for categorical prediction, M5Rules builds full trees instead of partially explored trees. Building partial trees leads to a greater

computational efficiency, and does not affect the size and accuracy of the resulting rules [23-25].

Random subspace

Random Subspace Method proposed by Ho [26]. The main idea of this technique is that, instead of using a single sub-space for classification, multiple subspaces are constructed from the original space by a random procedure. The Random Subspace technique produces a classifier using each random subspace. Then the final classification decision is achieved by combining all classification results. Although each classifier may not produce a good classification, their combination can be much better than using a single sub-space [27,28].

Regression by discretization

The regression by discretization approach allows the use of a classification algorithm in a regression task. It works as a pre-processing step in which the numeric target value is discretized into a set of intervals. A regression scheme that employs any classifier on a copy of the data that has the class attribute (equal-width) discretized. The predicted value is the expected value of the mean class value for each discretized interval (based on the predicted probabilities for each interval) [27-29].

RepTree

Quinlan [30] first introduced Reduced Error Pruning (REP) as a method to prune decision trees. REP is a simple pruning method, although it is sometimes considered to over prune the tree. A separate pruning dataset is required, which is considered a disadvantage of this method because data is normally scarce. However, REP can be extremely powerful when it is used with either a large number of examples or in combination with boosting. The pruning method that is used is the replacement of a subtree by a leaf representing the majority of all examples reaching it in the pruning set. This replacement is done if this modification reduces the error, i.e., if the new tree would give an equal or fewer number of misclassifications [27-31].

Results and Discussion

Experimental results

We used a Linear Regression model to predict characteristics of patients with an increased level of inflammation and compared the results with some classification algorithms (Decision table, Kstar, M5rules, Random Subspace, Regression by discretization and RepTree). A comparison of these techniques was presented. WEKA 3.6.8 software was used for analysis. WEKA is a collection of machine learning algorithms for data mining tasks and it is open source software. The software contains tools for data pre-processing, classification, regression, clustering, association rules and visualization [32,33].

Linear regression method was applied on our dataset and following regression function of the variables was generated:

CRP =

-0.0867 * Age +

0.2304 * BMI +

1.6537 * Analg=No +

1.8415 * OSP=No +

-0.287 * MMS +

.016 * EBV +

-0.1557 * HB +

58.4775 * HTC +

3.5499 * ALFA1 +

0.0062 * PRL +

0.029 * ANA +

-8.1658

The above result presents the obtained coefficients of the variables representing the regression output. The "+" signs denote positive correlations, whereas "-" signs denote negative correlations for the variables in the function [20,34].

Results evaluation

Before proceeding to final deployment of the model built by the data analyst, it is important to more thoroughly evaluate the model and review the model's construction to be certain it properly achieves the business objectives. Here it is critical to determine whether some important business issue has been insufficiently considered. At the end of this phase, the project leader should then decide exactly how to use the data mining results. The key steps here are the evaluation of results, the process review and the determination of next steps [23]. In our study, the comparison of the results obtained from several classification methods is shown in Table 4. The comparison parameters are the correlation coefficient (R^2-value), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and can be written as follows:

$$R^2 = 1 - \frac{\sum_{m=1}^{n}(t_{m,m} - y_{p,m})^2}{\sum_{m=1}^{n}(t_{m,m} - \bar{t}_{m,m})^2}$$

$$MAE = \frac{1}{n}\sum_{m=1}^{n}\left|y_{p,m} - t_{m,m}\right|$$

$$RMSE = \sqrt{\frac{\sum_{m=1}^{n}(y_{p,m} - t_{m,m})^2}{n}}$$

Where n is the number of data patterns, $y_{p,m}$ indicates the predicted, $t_{m,m}$ is the measured value of one data point m and $\bar{t}_{m,m}$ is the mean value of all measure data points [23].

Diagnosis/medication used	Positive cases N (%)
Hypertension	77 (82.80)
Diabetes mellitus or impaired glucose tolerance	34 (36.55)
Cardiovascular diseases	24 (25.81)
Gastroduodenal disorders	32 (34.41)
Chronic urinary tract disorders	51 (55.43)
Chronic obstructive pulmonary disease	13 (13.98)
Allergy (Rhinitis and/or Asthma)	9 (9.68)
Drugs allergy	18 (19.35)
Therapy with analgetics/non-steroidal anti-inflammatory drugs indicating patients with severe osteo-arthritis	23 (25.00)
Chronic skin disorders	51 (54.83)
Malignancy, including skin malignancy	14 (15.05)
Osteoporosis	30 (40.00)
Neuropsychiatric disorders	53 (57.00)
Cognitive impairment - MMS <24	30 (32.25)

Table 4: Proportion of subjects with diagnoses of chronic diseases or who use certain medications.

As can be seen from Table 5, Regression by discretization algorithm generated good results. This algorithm obtained a high correlation coefficient of 0.2441. Also, the algorithm has lower error rates among the other algorithms with a mean absolute error of 2.0757 and a root mean square error of 3.8528.

Discussion (Explanation of the Results in Light of the Current Knowledge)

The Linear Regression method, applied on the dataset that comprehensively describes the health status of the group of older persons with multiple morbid conditions, extracted a set of 11 variables, relevant enough to allow a significant model generation, in order to predict characteristics of persons with increased level of inflammation. A relatively large number of variables selected in the model are indicative of inflammation as a multifunctional disorder. This is not unexpected, since inflammation is known as an intermediate mechanism which links many interacting components, from the genetic and molecular to the phenotypic levels, into the common functional network [35]. The second explanation is that these results are reflective of a great inter-individual variability in clinical characteristics of older persons with increased level of inflammation. This explanation is in line with the recent studies of aging showing the aging phenotype, not as an uniform pattern, but as a heterogeneous mosaic of aging traits, which is a result of the complex interactions of environmental, genetic and stochastic factors [36]. In the same line, the studies of the frailty syndrome, the age-related phenotype characterised with weakness and motor dysfunction, showed the association of this syndrome with the physiological multisystem dysregulation, that means that the likelihood of frailty increases with increasing number of dysregulated physiological systems [37].

According to these studies, aging is considered as the complex remodeling process, that means that there is not a simple linear decline in the physiological functions over time, but instead, compensatory reactions follow detrimental changes, while the loss of some functions is maintained together with the over activation of some others [35,36]. Successful modeling of age-related functions considers the existence of efficient defense mechanisms. In this regard, chronic low grade inflammation is thought to be a result of the imbalance between inflammatory and anti-inflammatory regulatory networks and, as such, as the main driving force for age-related diseases development [35-38]. In other words, inflammation can be considered as a reflective of unsuccessful modeling, thereby linking the aging process with increased morbidity and mortality rates. The studies also say that when this point in the course of human life has passed, the mortality curve starts to slowdown instead of to continuously rise, reflecting the positive selection of very old individuals (85 and over) who are characterised with lower level of inflammation and better coping mechanisms, than their younger counterparts loaded with chronic aging diseases [38,39]. All this discussion was needed to help us explaining the obtained results on the negative association between the variable Age and the

Classification Algorithm	Correlation Coefficient	Mean Absolute Error	Root Mean Sqaure Error
Linear Regression	-0.0403	4.1321	5.7224
Decision Table	-0.0525	2.0679	3.9401
Kstar	-0.0041	2.4566	4.7958
M5rules	0.1904	2.1697	3.9943
Random Subspace	0.0649	1.9277	3.4506
Regression by discretization	0.2441	2.0757	3.8528
RepTree	-0.2978	1.9639	3.4934

Table 5: Comparison of the results of classification algorithms.

target variable CRP. Namely, this result might be due to the fact that the median age of the subjects in this sample was 69 years, indicating the age period when the disease burden is the largest. Obviously, the statement: the bigger the number of years of life, the higher inflammation level, does not fairly deal with our results, thereby supporting the remodeling theories of aging. The answer to the question why the variable sex (male/female) maintained unselected in the model, might not be due the lack of its relevance for the pathophysiology of inflammation, but rather the reason might be due to the mixed sample, or the insufficient sensitivity of the input variables for the differences in gender. In this regard, evidence suggests that women and men use different pathways in achieving age-related morbidity changes, or in longevity. In comparison to men, women are less dependent on genetics and more on environmental and societal factors [40].

Selection of the majority of other variables in this model also argues towards the proof of the conceptual framework of the remodeling theory of aging. According to this theory, the three main physiological systems, the metabolic, the neuroendocrine and the immune systems, are interconnected and capable of influencing each other during the age-related remodeling [35,38,41,42]. Chronic inflammation has been recognised as the common mechanism which links deteriorative changes in these systems together, accelerating the development of age-related diseases [38]. Elements of these systems and their relationships with chronic inflammation (indicated by the variable CRP) are likely to be captured by the selected variables BMI, MMS, PRL, EBV and ANA. Explanations for these variables are provided in the sections below.

In regard to the variable BMI, this is an anthropometric measure and well accepted marker of obesity. Obesity is known inflammatory condition, due to high secretory activity of the adipose tissue, that may promote insulin resistance (impaired insulin-mediated glucose metabolism), subsequently leading to glucose intolerance and diabetes, both disorders known as the major CV risk factors [43]. Among many secretory products with the origin in the adipose tissue, the main role in the pathogenesis of insulin resistance is prescribed to the inflammatory cytokines TNF-α and IL-6. Insulin resistance is considered to be the key mechanism of the insulin resistance (the metabolic) syndrome, clinically expressed by the cluster of several metabolic CV risk factors, the frequency of which increases with age, including: obesity, hypertension, impaired glucose tolerance, hyperinsulinemia and dyslipidemia [7]. Some other factors have subsequently been added to the syndrome, including impaired coagulation/fibrinolysis, increased blood viscosity and low grade chronic inflammation [44,45]. In this context we can consider the variables HB and HTC, selected in the regression model. Namely, these variables have already been identified as blood viscosity measures and, thus, as the part of the metabolic syndrome [45]. Specifically, the variable HB, indicating anemia, can also be viewed from the perspective of the frailty syndrome, which in older people is characterised by muscle wasting and low anthropometric performances [10]. Related to this, it has already been recognised that impaired body composition, not just obesity, but also the imbalance between the fat and the muscle contents, may influence the level of inflammation. Older age is associated with changes in both, in the level of inflammation and in body composition [46,47]. It is not surprisingly then that chronic inflammation has been recognised as the main reason why the frailty syndrome, despite the weight loss, increases the risk for the development of CV disease [48]. Based on this discussion, it is possible that, in addition to the variable HB, the variable BMI also has a dual function, indicating physical weakness and weight loss, rather than (or in addition to) the state of being overweight. The general implications of these results would be that any state associated with great deviations in body compositions (including either malnourishment or overweight)

can also increase the level of inflammation. This statement once again argues towards the remodeling theory of aging, indicating that there is the evolutionary necessity for the close relationships between the metabolic and the immune (inflammation-related) functions [49].

The selected variable ALFA1 indicates the increased hepatic synthesis of acute-phase proteins and in this context can be used as a complementary one to the variable CRP, that is, as an marker of inflammation [50]. We can speculate on the reasons why some other variables that indicate increased inflammation and have been used as the model`s input, including: LE, MO, LY, NEU and ALB, have not been not selected in the model. The reasonable explanation is that the inflammatory cytokine IL-6, known to be in the background of the increased synthesis of CRP, taken as the target variable in this model, can trigger a particular part of the inflammatory response [50].

According to the large body of evidence, there are multiple connections between the immune, the endocrine and the central nervous systems (CNS) [51]. The CNS regulates the immune system through neuroendocrine pathways and the autonomic nervous system, while the immune system sends signals to the brain via humoral factors, mainly cytokines, but also other immune mediators. In this terms, many cytokines and their receptors have been identified in the brain. Reversely, receptors for hormones, neurotransmitters and other neuropeptide mediators can be found on lymphocytes and other immune cells [51,52]. Knowledge on these communications at the molecular and cellular levels helps us understanding the observed clinically sound associations between inflammation and the immune mediated diseases, on the one hand, and the endocrine system related and mood disorders related diseases, on the other hand [51]. In this context, it has been known for a long time that the main neuroendocrine pathway which regulates the immune system functions is the hypothalamus-pituitary-adrenal (HPA) stress axis [51]. Reversely, the most powerful stimuli of this axis, in humans, is likely to be the cytokine IL-6, an inductor of acute-phase proteins synthesis, including also CRP, the main clinical marker of low grade inflammation [2]. Recent evidence indicates even more complex regulatory network to exist, including the neuroendocrine stress axis, obesity, inflammation and insulin resistance, linking the neuroendocrine, immune/inflammatory and metabolic pathways, into the common network [49]. In the same line of evidence, cognitive function decline (reflective of the brain degenerative/inflammatory processes) has been recognised as the part of the frailty syndrome (reflective of chronic inflammation and changes in body composition) [53]. Within this complex context, described above, we can find explanations for the selected variables MMS and PRL. The variable MMS indicates cognitive function decline, proposed to be the result of the brain aging and neurodegenerative brain disorders [54]. Evidence indicating that increased level of systems inflammation, together with metabolic and vascular changes, due to insulin resistance, are proposed triggering/accelerating factors in the pathogenesis of neurodegenerative disorders, can explain this observed link between the variable MMS and the target variable CRP. Another selected variable, PRL, indicating hormone prolactin serum concentration variations, might be also considered as the marker of the brain aging and degenerative brain disorders, elbeit focus shifted towards the pituitary gland, the central point of neuroendocrine control [55]. In line with the previous discussion is also the evidence indicating an association between the increased serum prolactin concentrations, insulin resistance and increased risks for CV disease development [56]. In fact, evidence is still insufficient on whether there is a pro-inflammatory or anti-inflammatory role of this hormone in the pathophysiology network of neuroendocrine and metabolic age remodelling, or whether there is increase, or decrease in this hormone

serum concentrations [57]. Our results also support this common notion that there is the complex remodeling process along with the course of age-related disease development, rather than simple linear decline in functions.

The selection of the variables EBV and ANA in the predictive model of inflammation can also be explained in the light of existing evidence and it is associated with the immuno-regulatory role of inflammation [2]. In this regard, one line of evidence indicates the role of cytokine IL-6 in enhancing humoral immunity, according to its described function as a stimulator of B-lymphocytes proliferation/maturation and immunoglobulins (antibodies) production [8]. The second line of evidence emphasises the role of low grade inflammation in accelerating aging of the immune system (immunosenescence). This happens in conditions of aging and the presence of age-related diseases. The process is characterised by the failure of specific immunity, reactivation of latent infections (in our sample indicated by the selected variable EBV) and spontaneous production of autoantibodies, even in the absence of autoimmune diseases (indicated by the selected variable ANA) [58]. In this context, ANA is a routinely used diagnostic marker in rheumatic autoimmune diseases, but its serum concentrations also increase spontaneously with age and are found to be elevated, as well in healthy elderly people, as in associations with different chronic diseases [59,60]. We can only speculate on the reasons why the variable EBV, indicating latent EBV infection, was selected in the model, but not other three variables (CMV, HBG and HPA) that indicate latent cytomegalovirus and Helicobacter pylori infections. According to the available evidence, the reason might be the synergistic action between EBV infection and chronic low grade inflammation in B lymphocytes stimulation [58].

Apart from these well known relationships involving chronic low grade inflammation, described above, our results also add new values to the knowledge of inflammation, by indicating the involvement of the locomotor system chronic inflammatory disorders, in this complex pathophisiology network. This is indicated by the two nominal variables selected in the model, Analg=No and OSP=No. Related to these findings, osteoporosis, indicated with the variable OSP, is a condition characterised with increased bone fragility and risks for fracture, due to enhanced bone resorption and/or decreased formation. This is a well known inflammation-mediated aging disease which predominantly affects women [2,18]. In fact, multiple pathogenetic mechanisms have been proposed for this disease, including genetic, biomechanical, metabolic, hormonal and inflammatory factors [18]. We can speculate that the option "OSP=No" of this selected variable does not mean the negative selection of subjects free from this disease, but on the contrary, implicates early stages of this disease development, when the degree of inflammation, produced locally in bone tissue, is also likely to be high. These early stages are complementary to the diagnosis of osteopenia (the condition in which the bone tissue content is reduced, but at lesser degrees than in osteoporosis). In contribution to this statement is the fact that osteopenia has been recognised as the component of the frailty syndrome, known to be an inflammatory-mediated condition [10]. Once when this mentioned bone disease turns into its advanced stage, the fairly large bone loss is unable to elicit any inflammatory reaction. Similar explanation might state for the second variable, "Analg=No", indicating older people who does not use analgesics and non-steroidal anti-inflammatory drugs. This statement is likely to exclude patients with severe osthteo-arthritis (who does use these drugs), allowing for those ones with early stages of this disease. In this regard, recent evidence suggests that osteo-arthritis, a condition characterised with the fragility and loss of cartilage tissue in the weight-bearing joints, including prevalently the knees and the hips, has been recognised as an

inflammation-mediated disorder and important source of inflammatory mediators [61]. Systemic inflammation, associated with this condition, affects the susceptibility of these patients to the development of chronic aging diseases, notably atherosclerotic CV and Alzheimer`s disease. The problem with this provided explanation for the variable "Analg=No" is the lack of the classification of osteo-arthritis disease, which makes the relationships between the disease activity and the level of inflammation difficult. Additional thing is that use of analgesics can mask the inflammation state. This all enables only general statement upon the selection of this variable, indicating analgesics use (or non use), and the existence of the conditions complementary to osteo-arthritis, without taking into account the exact position of this variable in the ordinal scale values, by means of altering the valus "yes" or "no").

Conclusions

We generated the model to predict features of older people with low grade inflammation, by using a large set of simple health data collected in a way to determine the health status of patients from many aspects, systematically. The model is likely to capture all main relationships known to link the IL-6 gene expression (a hallmark of inflammation) with the metabolic, the neuroendocrine and the immune system pathways, implicating inflammation as a part of the highly conserved pathophysiology network. Apart from these core elements, some new components of the network have been emphasised, including inflammation-mediated bone and joint disease and particular neuroendocrine pathways, with the pituitary hormone prolactin being involved in. When the existing knowledge was used to explain these variables selection, some interesting facts have emerged, leading to new hypotheses generation. One hypothesis states that there is an association between increased level of inflammation and changes in body`s composition, which might be the main driving mechanism of this network. Another hypothesis deals with the notion that there is a nonlinear remodeling of the metabolic, the immune and the neuroendocrine pathways during aging and age-related disease development. Whether the components and pathways outlined in this model are of universal importance, or whether they just reflect the characteristics of the study sample, have to be explored by repeating the same procedure on other samples. This presented approach based on using systematically collected, easily available health parameters, can be used to support decision-making when planning preventive strategies, to initially mapping components when an unexplored problem is to be solved, or to supplement some other system`s methods. Even more important, this approach could be useful to seeking for simple, low-cost markers, for making prediction in the area of chronic aging disease. According to our results, parameters indicating analgesics use (variable Analg), increased blood viscosity (variable HTC) and acute-phase response (variable ALFA1), deserve such attention.

References

1. Medzhitov R (2010) Inflammation 2010: New adventures of an old flame. Cell 140: 771-776.

2. Papanicolaou DA, Wilder RL, Manolagas SC, Chrousos GP (1998) The pathophysiologic roles of interleukin-6 in human disease. Ann Intern Med 128: 127-137.

3. Ridker PM, Hennekens CH, Buring JE, Rifai N (2000) C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. N Engl J Med 342: 836-843.

4. Koenig W, Sund M, Frohlich M (1999) C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results from the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease). Circulation 99: 237-243.

5. Harris TB, Ferrucci L, Tracy RP, Corti MC, Wacholder S, et al. (1999)

Associations of elevated interleukin-6 and C-reactive protein levels with mortality in the elderly. Am J Med 106: 506-512.

6. Festa A, D'Agostino R Jr, Howard G, Mykkanen L, Tracy RP, et al. (2000) Chronic subclinical inflammation as part of the insulin resistance syndrome: the Insulin Resistance Atherosclerosis Study (IRAS). Circulation 102: 42-47.

7. DeFronzo RA, Ferrannini E (1991) Insulin resistance. A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. Diabetes Care 14: 173-194.

8. Ershler WB (1993) Interleukin-6: a cytokine for gerontologists. J Am Geriatr Soc 41: 176-181.

9. Nyugen J, Agrawal S, Gollapudi S, Gupta S (2010) Impaired functions of peripheral blood monocyte subpopulations in aged humans. J Clin Immunol 30: 806-813.

10. Ershler WB, Keller ET (2000) Age-associated increased interleukin-6 gene expression, late-life diseases, and frailty. Annu Rev Med 51: 245-270.

11. Soerensen M, Dato S, Tan Q, Thinggaard M, Kleindorp R, et al. (2013) Evidence from case-control and longitudinal studies supports associations of genetic variation in APOE, CETP, and IL6 with human longevity. Age (Dordr) 35: 487-500.

12. Dahlstrom O, Timpka T, Hass U, Skogh T, Thyberg I (2008) A simple method for heuristic modeling of expert knowledge in chronic disease: identification of prognostic subgroups in rheumatology. Stud Health Technol Inform 136: 157-162.

13. Alley DE, Seeman TE, Ki Kim J, Karlamangla A, Hu P, et al. (2006) Socioeconomic status and C-reactive protein levels in the US population: NHANES IV. Brain Behav Immun 20: 498-504.

14. Ebling B, Majnaric-Trtica L, Gmajnic R, Ebling Z, Vranjes Z (2007) Psycho-social aspects of measures aimed at decreasing prevalence of chronic diseases in the population of returnees in the Osijek Region, Croatia. Coll Antropol 31: 441-450.

15. Trtica-Majnaric L, Zekic-Susac M, Sarlija N, Vitale B (2010) Prediction of influenza vaccination outcome by neural networks and logistic regression. J Biomed Inform 43: 774-781.

16. Majnaric-Trtica L, Vitale B, Martinis M, Reiner Z (2010) A view at the future--a dynamical, protocol-based and computationally intensive approach in cardiovascular risk assessment. Coll Antropol 34: 437-445.

17. Majnaric Lj, Zekic-Susac M (2012) Elucidating clinical context of lymphopenia by nonlinear modeling. Expert Systems with Applications 39: 10889-10897.

18. Soltow QA, Jones DP, Promislow DE (2010) A network perspective on metabolism and aging. Integr Comp Biol 50: 844-854.

19. Schneider A, Hommel G, Blettner M (2010) Linear regression analysis: part 14 of a series on evaluation of scientific publications. Dtsch Arztebl Int 107: 776-782.

20. Mihaescu MC (2011) Classification of learners using linear regression. Proceedings of the Federated Conference on Computer Science and Information Systems Sep 18-2,201, Szczecin Poland 717-721.

21. Kohavi R, John GH (1997) Wrappers for feature subset selection. Artificial Intelligence 97: 273–324.

22. Cunnigham SJ, Holmes G (1999) Developing innovative application in agriculture using data mining. In proceedings of the Southeast Asia regional computer confederation conference.

23. Küçüksille EU, Selbas R, Sencan A (2011) Prediction of thermodynamic properties of refrigerants using data mining. Energy conversion and management 52: 836-848.

24. Wou S, Zhao X, Shao H, Ren D (2004) Cold rolling process data analysis based on svm. In proceedings of the third international conference on machine learning and cybernetics. Shangai 2: 792 - 795.

25. Goodwin L, VanDyne M, Lin S, Talbert S (2003) Data mining issues and opportunities for building nursing knowledge. J Biomed Inform 36: 379-388.

26. Ho KT (1998) The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20: 832-844.

27. Nisanci MH, Küçüksille EU, Cengiz Y, Orlandi A, Duffy A (2011) The prediction of the electric field level in the reverberation chamber depending on position of stirrer. Expert Systems with Applications 38: 1689-1696.

28. Nguyen N, Liu W, Venkatesh S (2007) Random subspace two-dimensional pca for face recognition. Lecture Notes in computer Science 4810: 655-664.

29. Leon F, Cristian B, Zaharia MH (2006) Load balancing in mobile intelligent agents framework using data mining classification techniques. In Eighth international conference on development and application systems, 25-27 May, Romania, 459-466.

30. Quinlan JR (1992) Learning with continuous classes. Proceedings of the Australian joint conference on artificial intelligence 343-348.

31. Licamele K, Getoor L (2006) Predicting protein-protein interactions using relational features. In Proceedings of ICML workshop on statistical network analysis.

32. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, et al. (2009) The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter 11: 10-18.

33. WEKA: Weka 3: Data Mining Software in Java.

34. Siew HL, Nordin J (2012) Regression techniques for the prediction of stock price trend, Statistics in Science, Business and Engineering (ICSSBE) Langkawi.

35. Franceschi C, Valensin S, Bonafe M, Paolisso G, Yashin AI, et al. (2000) The network and the remodeling theories of aging: historical background and new perspectives. Exp Gerontol 35: 879-896.

36. Cevenini E, Invidia L, Lescai F, Salvioli S, Tieri P, et al. (2008) Human models of aging and longevity. Expert Opin Biol Ther 8: 1393-1405.

37. Fried LP, Xue QL, Cappola AR, Ferrucci L, Chaves P, et al. (2009) Nonlinear multisystem physiological dysregulation associated with frailty in older women: implications for etiology and treatment. J Gerontol A Biol Sci Med Sci 64A: 1049-1057.

38. Vasto S, Candore G, Balistreri CR, Caruso M, Colonna-Romano G, et al. (2007) Inflammatory networks in ageing, age-related diseases and longevity. Mech Ageing Dev 128: 83-91.

39. Manton KG (1999) Dynamic paradigms for human mortality and aging. J Gerontol A Biol Sci Med Sci 54: B247-254.

40. Franceschi C, Motta L, Valensin S, Rapisarda R, Franzone A, et al. (2000) Do men and women follow different trajectories to reach extreme longevity? Italian Multicenter Study on Centenarians (IMUSCE). Aging (Milano) 12: 77-84.

41. Paolisso G, Barbieri M, Bonafe M, Franceschi C (2000) Metabolic age modelling: the lesson from centenarians. Eur J Clin Invest 30: 888-894.

42. Franceschi C, Bonafe M, Valensin S (2000) Human immunosenescence: the prevailing of innate immunity, the failing of clonotypic immunity, and the filling of immunological space. Vaccine 18: 1717-1720.

43. Bastard JP, Maachi M, Lagathu C, Kim MJ, Caron M, et al. (2006) Recent advances in the relationship between obesity, inflammation, and insulin resistance. Eur Cytokine Netw 17: 4-12.

44. Sakkinen PA, Wahl P, Cushman M, Lewis MR, Tracy RP (2000) Clustering of procoagulation, inflammation, and fibrinolysis variables with metabolic factors in insulin resistance syndrome. Am J Epidemiol 152: 897-907.

45. Meiselman HJ (1999) Hemorheologic alterations in hypertension: chicken or egg? Clin Hemorheol Microcirc 21: 195-200.

46. FitzGerald L, M Macey P, Brecht ML (2012) Pathways to interleukin-6 in healthy males and serious leisure male athletes: physical activity, body composition and age. PLoS One 7: e40513.

47. Pedersen BK, Febbraio M (2005) Muscle-derived interleukin-6--a possible link between skeletal muscle, adipose tissue, liver, and brain. Brain Behav Immun 19: 371-376.

48. Phan HM, Alpert JS, Fain M (2008) Frailty, inflammation, and cardiovascular disease: evidence of a connection. Am J Geriatr Cardiol 17: 101-107.

49. Wellen KE, Hotamisligil GS (2005) Inflammation, stress, and diabetes. J Clin Invest 115: 1111-1119.

50. Ramadori G, Christ B (1999) Cytokines and the hepatic acute-phase response. Semin Liver Dis 19: 141-155.

51. Marques-Deak A, Cizza G, Sternberg E (2005) Brain-immune interactions and disease susceptibility. Mol Psychiatry 10: 239-250.

52. Dorshkind K, Horseman ND (2000) The roles of prolactin, growth hormone, insulin-like growth factor-1 anf thyroid hormones in lymphocyte development

and function: insights from genetic models of hormone and hormone receptor deficiency. Endocine Review 21: 202-312.

53. Avila-Funes JA, Amieva H, Barberger-Gateau P, Le Goff M, Raoux N, et al. (2009) Cognitive impairment improves the predictive validity of the phenotype of frailty for adverse health outcomes: the three-city study. J Am Geriatr Soc 57: 453-461.

54. Marlatt MW, Lucassen PJ, Perry G, Smith MA, Zhu X (2008) Alzheimer's disease: cerebrovascular dysfunction, oxidative stress, and advanced clinical therapies. J Alzheimers Dis 15: 199-210.

55. Ferrari E, Cravello L, Muzzoni B, Casarotti D, Paltro M, et al. (2001) Age-related changes of the hypothalamic-pituitary-adrenal axis: pathophysiological correlates. Eur J Endocrinol 144: 319-329.

56. Harati H, Hadaegh F, Saadat N, Azizi F (2009) Population-based incidence of Type 2 diabetes and its associated risk factors: results from a six-year cohort study in Iran. BMC Public Health 9: 186.

57. Barzilai N, Gabriely I, Atzmon G, Suh Y, Rothenberg D, et al. (2010) Genetic studies reveal the role of the endocrine and metabolic systems in aging. J Clin Endocrinol Metab 95: 4493-4500.

58. Copie-Bergman C, Niedobitek G, Mangham DC, Selves J, Baloch K, et al. (1997) Epstein-Barr virus in B-cell lymphomas associated with chronic suppurative inflammation. J Pathol 183: 287-292.

59. Satoh M, Chan EK, Ho LA, Rose KM, Parks CG, et al. (2012) Prevalence and sociodemographic correlates of antinuclear antibodies in the United States. Arthritis Rheum 64: 2319-2327.

60. Grainger DJ, Bethell HW (2002) High titres of serum antinuclear antibodies, mostly directed against nucleolar antigens, are associated with the presence of coronary atherosclerosis. Ann Rheum Dis 61: 110-114.

61. Berenbaum F (2013) Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). Osteoarthritis Cartilage 21: 16-21.