Integrated Computer-Aided Engineering 19 (2012) 151–163
DOI 10.3233/ICA-2011-0386
IOS Press

A model for mining material properties for radiation shielding

Hisham Al-Mubaid* and Dawood Moazzam
Department of Computer Science, University of Houston-Clear Lake, Houston, TX, USA

Abstract. Radiation shielding has been an active subject of research in the space industry for many years. We propose a new model for mining material properties for radiation shielding. This work represents an effective way of using learning and feature selection for selecting the material properties that most affect the shielding effectiveness of materials. The methodology relies on machine learning as a measure for the identified subsets of material properties for radiation shielding. This is a new direction in working with radiation shielding using purely computational techniques with machine learning. The experimental results showed that the approach is quite effective in eliminating redundant features and identifying the most significant properties related to radiation shielding capability of materials. For example, we have identified some material properties, besides *Density*, like *Heat of Fusion, Atomic Number*, *X-ray Absorption Edge*, *Electrical Resistivity*, and *Specific Heat Capacity* that are highly related to the radiation shielding as they have been proved computationally. The evaluation results also show that all machine learning algorithms can induce more robust separating models for the subsets of reduced number of features that are highly significant in the domain of radiation shielding.

1. Introduction

With the advances in material science and engineering, it is becoming more interesting and useful to work with computational techniques into material data and properties for certain application areas like radiation shielding. In computational material science, computational techniques, including data mining techniques, are used to expand our knowledge and findings in material properties and behavior [23]. The properties of materials (e.g. density, melting point, specific heat capacity, linear CTE, thermal conductivity, etc.) are useful and important for many domains and application areas including industrial engineering, space science, mechanical engineering, and manufacturing engineering. Studying material properties, from data mining perspective, supports more progress in these application domains [1,3,5,13]. For example, in manufacturing industries, the manufacturer may be interested in finding the most important material properties that make steel, aluminum, and alloy have the similar degrees of thermal conductivity [41,42]; or the most significant properties that make aluminum and alloy have comparable ranges of linear CTE (*coefficient of thermal expansion*) [3]. Another example, in space science, it is vital to protect human in space from all types of space radiation. Researchers in space science are working actively on how to reduce and block space radiation due to the harmful effect of energy loaded with the ions which comes from the energy in transit that exists in the form of high-speed particles and electromagnetic waves [33]. Moreover, in the past decade, NASA has invested significantly in research and studies on shielding from space radiation. For example, in the CEV project, besides the load and weight of the vehicle, the radiation exposure is highly important research parameter [http://www.spaceref.com/]. Generally, radiation can be classified as: (1) ionizing radiation and (2) non-ionizing radiation. The space radiation typically consists of high-energy charged particles comprising what have come to be known as *ionizing radiation*. The *ionizing radiation* contains high levels of energy that is enough to remove electrons from the orbits of atoms and thus resulting in charged particles [33,40]. There are three naturally occurring sources of space radiation: trapped radiation, galactic cosmic radiation (GCR), and solar particle events (SPE) [40,47]. Space shuttles and vehicles typically are built with materials having high

*Corresponding author: Dr. Hisham Al-Mubaid, Chair, Computer Information Systems. Associate professor, Computer Science and Computer Information Systems, University of Houston-Clear Lake, 2700 Bay Area Blvd, Box 40, Houston, TX 77058, USA. Tel.: +1 281 283 3802; E-mail: hisham@uhcl.edu.

ISSN 1069-2509/12/$27.50

rates of shielding from space radiation. So little has been known about the key constituents of shielding goodness in certain materials, e.g. *hydrogen*. For example, it has been known that hydrogen atoms are good in absorbing and dispersing radiation. But what are the other attributes and constituents of hydrogen and other materials that affect shielding capabilities. Data mining methods can be effective in discovering and extracting useful patterns from data and organizing the data for efficient retrieval, prediction, clustering, and classification applications [1,21,23]. In this paper, we want to study and analyze the *shielding goodness* of materials by a new way of modeling material properties with feature selection and machine learning [8,16,25, 44]. For that, we propose a methodology for estimating radiation shielding from the properties of the materials. This work uses the advances in feature selection with learning and classification [26,49] to discover the most important properties of the materials in relation to radiation shielding.

Contribution: the main contributions of this work are (1) a new model for mining material properties to determine the most significant properties for radiation shielding, and (2) a new perspective of addressing the problem of radiation shielding: to the best of our knowledge this is the first work that handles this problem using purely computational techniques. The material properties that have been found to be related with radiation shielding have been verified in the literature. Moreover, this work represents a new direction in working with radiation shielding using purely computation techniques with data mining and machine learning [16,17,20,26,48].

We know density is the most common property related to shielding capability of materials (i.e., material with higher density tends to be better for radiation shielding) [7]. For example, it has been known, for decades, that *lead Pb* is one of the best materials (Table 1) for radiation shielding due to its high *density* (11.34 g/cc; see Table 3). It has been used in X-ray shielding vests in the medical domain for long time. But, the question is, can we identify other materials with low density, and thus lower weight, that still have good shielding rate? In other words, what other features, besides density, can help in enhancing the shielding effectiveness of materials? And, in general, what are the material properties that influence the radiation shielding capabilities of materials the most?

The experiments and evaluation results proved that (1) the new techniques is quite effective in reducing the number of significant features to 30, 15, or 5 and removing all the non-significant and redundant features. Therefore, it represents an effective methodology for identifying the most significant features that are affecting the radiation shielding capability of materials; and (2) the material databases and material properties can be easily and feasibly analyzed and utilized in data mining and the outcomes can be useful and practical for the space industry and other industrial material science and engineering applications. For example, the method was able to identify some material properties, besides *density*, like *Heat of Fusion, Atomic Number*, *X-ray Absorption Edge*, *Electrical Resistivity*, and *Specific Heat Capacity* that are highly related to, and can improve, the radiation shielding capability of materials.

Table 1
Sample of materials

1040 Aluminum	Holmium, Ho
2195 Aluminum Composition	Hydrogen, H2
Spec	INCONEL Alloy 600
301 Stainless Steel	Indium, In
314 Stainless Steel, annealed plate	Iodine, I2
92% Alumina, opaque	Iridium, Ir; Cold-Drawn
96% Silica Glass	Iron (II, III) Oxide, Fe3O4
Actinium, Ac	(Magnetite)
Alclad Aluminum 6061-T6, T651	Iron, Fe
Alclad Aluminum 7075-T6, T651	Lanthanum, La
Aluminum 2024-T6	Mercury, Hg
Aluminum 2090-T86	Neodymium, Nd
Helium, He	Oxygen, O2
Hogen Duralloy H100M	Ozone, O3
Krypton, Kr	Osmium, Os; Arc-Melted Sil-
Molybdenum	ver, Ag
Lithium, Li	Silver, CP Grade
Lutetium, Lu	Silver, CP Grade, 70% Cold
Lead (II) Oxide, PbO (Litharge)	Worked
Lead, Pb	Strontium, Sr
Lithium Aluminum Hydride,	Sulfur, S, Beta
LiAlH4 Magnesium, Mg;	Tantalum, Ta; Annealed
Neon, Ne	Osmium, Os; Cold-Worked
Nickel, Ni	
Nitrogen, N2	
Osmium, Os; Annealed	

The rest of this paper is organized as follows. The next section presents the methodology to identify the most significant features related to radiation shielding using feature selection and learning framework. Feature selection techniques are explained in Section 3. Section 4 includes the evaluation and discussion. Finally, Section 5 presents conclusions.

2. The methodology

In its simplest form, we can view the materials and their properties as a projection on the categories of radiation shielding of materials formed in the input dataset.

Table 2
A sample collection of the properties used in evaluation

Properties					
Basic elements (% components) properties			None basic element properties		
Actinium, Ac (%)	Cerium, Ce (%)	Gold, Au (%)	Adapter Temperature	Hardness, Shore D	Temperature, Inert (Â°C)
Aluminum, Al (%)	Cesium, Cs (%)	Hafnium, Hf (%)	Aging Temperature	Hardness, Vickers	Melt Flow (g/10 min)
Americium, Am (%)	Chlorine, Cl (%)	Helium, He (%)	Annealing Point (Â°C)	Haze (%)	Melting Point (Â°C)
Antimony, Sb (%)	Chromium, Cr (%)	Holmium, Ho (%)	Annealing Temperature (Â°C)	Heat of Formation (kJ/mol)	Middle Barrel Temperature (Â°C)
Argon, Ar (%)	Cobalt, Co (%)	Hydrogen, H (%)			
Arsenic, As (%)	Copper, Cu (%)	Indium, In (%)	Apparent Bulk Density (g/cc)	Heat of Fusion (J/g)	Minimum Service Temperature, Air
Astatine, At (%)	Curium, Cm (%)	Iodine, I (%)		Heat of Vaporization (J/g)	
Barium, Ba (%)	Dysprosium, Dy (%)	Iridium, Ir (%)	Hardness, Brinell		Modulus of Elasticity (GPa)
Berkelium, Bk (%)	Einsteinium, Es (%)	Iron, Fe (%)	Hardness, Knoop	Neck In (cm)	
Beryllium, Be (%)	Erbium, Er (%)	Krypton, Kr (%)	Hardness, Mohs	Notched Tensile Strength (MPa)	Moisture Content
Bismuth, Bi (%)	Europium, Eu (%)	Lanthanum, La (%)	Hardness, Rockwell A		Molecular Weight (g/mol)
Boron, B (%)	Fermium, Fm (%)	Lead, Pb (%)	Hardness, Rockwell B	Nuclear Spin	Solubility (ppm)
Bromine, Br (%)	Flourine F (%)	Lithium, Li (%)	Hardness, Rockwell C	Oxidative Induction Time (OIT) (min)	Solution Temperature (Â°C)
Cadmium, Cd (%)	Francium, Fr (%)	Lutetium, Lu (%)	Hardness, Rockwell R		
Calcium, Ca (%)	Gadolinium, Gd (%)	Magnesium, Mg (%)	Capacity (J/g-Â°C)	Shear Modulus (GPa)	Specific Heat capacity
Californium, Cf (%)	Gallium, Ga (%)	Manganese, Mn (%)	Specific Surface Area (mÂ^2/g)	Shear Strength (MPa)	Mooney Viscosity
Carbon, C (%)	Germanium, Ge (%)	Mercury, Hg (%)		Solidus (Â°C)	
			Spectral Line 1		
Total: 102 properties			*Total: 148 properties*		

Let k be the number of categories for the goodness of radiation shielding. We define '*shielding goodness*' of a material m_i as the rate at which the material m_i can shield and protect from radiation. Initially, we start with only two categories, $k = 2$, *good shielding* and *bad shielding*. A dataset of material properties given consists of n materials, each material has p properties, and all properties are continuous real values {*note: we use the terms properties and features interchangeably*}. Let us represent the set M of materials as M = {m_1, m_2, ..., m_n}. We also represent the set F of all features as F = {f_1, f_2, ..., f_p}. Thus each material m_i can be represented as: $m_i \subseteq R^p$ and the material data can be viewed as an $[nxp]$ matrix T where $T \subseteq R^{nxp}$, and radiation shielding capability as a category or output parameter (see properties in Table 2). That is, $t_{ij} \in T$ represents the property j of material i. This model relies on representing the materials and their properties as a projection on the categories of radiation shielding of materials formed in the input dataset (as shown in Table 3). Let $k$be the number of categories for the *goodness* of radiation shielding. We define '*shielding goodness*' of a material m_i as the rate at which this material, m_i, can shield and block the radiation. Let C = {C_1, C_2, ..., C_k} be a finite set of classes such that each material ($i.e.$,each row in matrix T) is assigned to one of these classes. We would like to select a subset s of the set F {$s \subseteq F$} that includes the most *significant* features that are correlated with the radiation shielding capability of materials. Initially, we start with only two categories, $k = 2$ (*e.g. good shielding* and *bad shielding*).

Then, we want to be able to identify the salient and significant properties of materials that affect the shielding goodness and improve the goodness of radiation shielding. The correlation between the various properties of a material on the shielding goodness of that material is one of the important outcomes that we want to induce from the data. That is, we would like to identify those features that are highly correlated with the classes of shielding goodness of the materials, and since the classes represent different radiation shielding rates then the resulting features will represent those features that are highly correlated with the shielding capabilities of materials.

We collected the features of the materials from multiple material database sites on the web including the material web: www.matweb.com, Engineers Edge: www.EngineersEdge.com, and Metal Science Resource: http://www.lib.iastate.edu/collections/eresourc/matsci.html [10,31,46]. A sample list of some of the materials is included in Table 1 whereas Table 2 and Table 3 contain more samples of materials and properties that we collected for the materials. Table 3 contains data for 8 properties of a sample of 10 materials. We used software that computes the Linear Energy Transfer (LET) vs. range for materials (LET ver 1.24, *Brookhaven National Laboratory*) [22]. This software uses the beam energy to calculate the *LET* value by using the semi-empirical formulas developed by J. F. Ziegler et al. [52, 54]. Then we can compute for each material what will be known as *material range* (*aka* RMAT) in {g/cm^2} by multiplying the energy value calculated by the *LET*

Table 3
Sample of material data: a sample of 8 properties (columns) for 10 materials (rows)

Materials	Properties							
	Density: g/cc	Tensile Strength, Ultimate (MPa)	Tensile Strength, Yield (MPa)	CTE, linear (Aμm/m-A°C)	Specific Heat Capacity	Dielectric Constant	Material Range (g/cm^2)	Elongation at Break (%)
2195 Aluminum Composition Spec	2.71	122	205	0.3	0.3	0.9	2.878	20
Titanium Ti-6Al-4V (Grade 5), Annealed	4.43	950	880	9.20	0.5263	0.3	3.241	14
Aluminum 6082-T6	2.7	310	260	0.3	0.3	0.3	2.921	10
Helium, He	0.00016	0.3	0.3	0.3	5.193	0.3	2.264	0.3
Hydrogen, H	0.00008	0.3	0.3	0.3	14.2857	0.3	1.029	0.3
Lead, Pb	11.34	18	0.3	29.1	0.129	0.3	5.055	0.3
High Density Polyethylene (HDPE),	1.46	15.2	13.1	200	0.3	2.05	2.065	25.0
Water, H2O	0.99	0.3	0.3	69	4.1819	77.9 79.0	2.202	0.3
Aluminum, Al	2.69	0.3	0.3	25.5	0.9	0.3	2.92	0.3
Silicon Phosphide, SiP	0.3	0.3	0.3	0.3	0.3	0.3	2.346	0.3

Table 4
The datasets of material properties

Dataset	Contents		No. of Materials	No. of Features	Material Range (g/cm^2)
	Materials	Properties			
DS1	All materials	All properties	181	250	0.38–6.68
DS2	Only basic elements	All properties	120	250	0.38–5.31
DS3	Only Compound materials	All properties	61	250	0.75–6.68
DS4	All materials	only % component elements	181	102	0.38–6.68
DS5	All materials	properties only	181	148	0.38–6.68

Table 5
The distribution of materials into two classes (and five classes) based on the value of *material range* property

Classes		Material range	Number of materials	Total number of features
2-Class	C_1	0 to 3.7	85	250
	C_2	> 3.7	96	250
5-Class	C_1	< 2.9	38	250
	C_2	> 2.9 to 3.7	47	250
	C_3	> 3.7 to 4.5	24	250
	C_4	> 4.5 to 5.5	56	250
	C_5	$5.5 \leqslant$	16	250

by the density {in g/cm^3}. This *RMAT* is basically the energy range of a *50 MeV* (in this case) proton beam for different materials [35]. If the resulting RMAT value is below certain cutoff (for example 3.7) then it is considered good material for radiation shielding, otherwise, it is not a good material for radiation shielding [47,54]. For example, let us consider the *Aluminum* (*Al*); the density of *Al* is 2.699 g/cm^2 and the range at 50 *MeV* is 10760 μm; then the RMAT can be computed as follows: (10760×10^{-6}) m $\times$ 100 cm/m $\times$ 2.69 g/cm^3 $=$ 1.076 * 2.69 g/cm^2 $=$ 2.89 g/cm^2. Thus we associate with each material an RMAT value as a property of the shielding capability of that material [35] (we will use the term '*material range*' to refer to *RMAT*).

Classes of materials: We divided the materials into $k$classes based on the *radiation shielding* (*material range* property) of each material. That is, material having relatively low '*material range*' value ($<$ 3.7, for example) where grouped together in the same class as these materials are considered good for radiation shielding. This puts all materials into two classes with a *material range* threshold (3.7), used as cutoff point as shown in Table 5. In the aerospace science and applications, materials with LET material range $\leqslant$ 3.7 are considered good for radiation shielding [47,54].

We also divided the materials into five classes (also based on *material range* values) as shown in Table 5. The materials in every class, in Table 5, are considered similar in terms of their capability of radiation shielding. For example, the 24 materials in class C_3 are considered having the same level of radiation shielding because their *material range* values are in the same level, i.e., between 3.7 to 4.5 g/cm^2. Thus this clustering of radiation shielding values is finer grained than those in Table 5. Next, the computational model uses five machine learners (See Table 7): Naïve Bayes (NB), Decision Trees (J48), Random Forest, SVM (LibSVM), and Decision Table [6,37,50] to induce separation models that will be applied to estimate the accuracy of the identified significant feature sets [16,48]. We used 10-fold cross validation (10F-CV) for computing accuracy in each experiment [39]. We use the advances in feature selection methods to be able to select and identify the features that influence the radiation shielding capability of materials in each class. The feature selection techniques that are used in our work, explained in the next section, are independent of the learning algorithms.

2.1. The proposed model

To understand the relationship between the various properties of materials and the radiation shielding we will have to use feature selection and learning. We first explain the relation between feature set and learning-classification accuracy. The learning-classification accuracy is one of several methodologies for evaluating feature subsets [9]. We emphasize that the relationship between them is existing and meaningful. Other evaluation functions for feature subsets include data intrinsic measures and incremental error rate [9,21]. In the following we specify our model and present the problem formulation. The proposed model consists of a combination of feature selection, machine learning [51,17], and a semi-empirical model to compute the shielding effectiveness of every material as follows:

1. We divide all materials into two classes based on shielding effectiveness (good shielding and bad shielding); and we used a cutoff value for *RMAT* to determine the class of each material.
2. We use filtering feature selection to rank all features based on the radiation shielding classes of materials so as to extract reduced feature set.
3. Machine learners are employed to induce separating models using reduced sets of features and to estimate the accuracy with the reduced features.

We use five different feature selection techniques (explained in Section 3) and five learners with various setting of reduced feature sets including 3 to 30 features. Large number of combinations of feature selection, learner, and number of features have been examined and evaluated.

Table 6
Distribution of materials in the two classes for each of the five datasets

Dataset	C1	C2
DS1	85	96
DS2	45	75
DS3	40	21
DS4	85	96
DS5	85	96

Given a dataset DS_i of $n$items and dimension p; also given two classes C_1 and C_2 such that each item in DS_i is either in C_1 or C_2which implies that:

$$C_1 \cup C_2 = DS_i \text{ and } C_1 \cap C_2 = \emptyset.$$

Let F be the set of all dimensions (features) in DS_i, and let F^x and F^y be two subsets of F($F^x \subseteq F$ and $F^y \subseteq F$.)

Let $LCA(L_j, DS_i, F^x)$ be the learning-classification accuracy of learner L_j on dataset *DSi* based on the two classes (C_1, C_2) and using the feature subset F^x. The learner L_j induces a separating model that distinguished between the two classes C_1 and C_2 and tries to assign each data item into C_1 or C_2 based on this induced model. If

$$LCA(L_j, DS_i, F^x) > LCA(L_j, DS_i, F^y)$$

then the subset F^x is more correlated to the classes (C_1 and C_2) of DS_i than F^y in this task. This implies that the learner, L_j, was able to induce better separating model with the feature set F^x than with F^y.

If $F^x \subset F^y$ then the extra features in F^y are noise or redundant.

Since the two classes C_1 and C_2, in this research, are classes of radiation shielding (C_1 includes the materials with good shielding and C_2 contains materials with bad shielding capability) then the set F^x contains more significant features than F^y for radiation shielding.

We use the accuracy with all features, $LCA(L_j, DS_i, F)$, as a *baseline* and so, if a reduced dataset F^x is extracted from F (i.e. $F^x \subseteq F$) using a filtering feature selection technique then we estimate that

$LCA(L_j, DS_i, F^x) > LCA(L_j, DS_i, F)$ implies that the reduced feature set F^x includes more significant features and the learner can assign the items of DS_i into C_1 or C_2 using the reduced set of features F^x better than with all features F.

Table 7
Learning-classification accuracy with all features for each dataset and with every classifier. The materials in each dataset are categorized into two classes

Dataset	No. of classes	Machine Learner:				
		Naïve Bayes	LibSVM	Decision Table	Random Forest	J48
DS1	2	80.5	84.7	87.5	84.2	83.7
DS2	2	88.3	82.5	92.5	91.7	92.5
DS3	2	71.4	77.1	83.7	81.9	73.8
DS4	2	71.8	71.9	72.9	72.9	76.3
DS5	2	84.5	83.6	89.5	84.0	86.5

All features; 2 classes; No FS.

3. Feature selection

Feature selection methods usually are divided into two classes: feature wrapper methods and feature filtering methods [18]. In the wrapper methods, features are selected with respect to a given learning and classification algorithm to improve the accuracy of a given learner with reduced number of features. The filtering methods, on the other hand, attempt to select a subset of features independent of any learning or classification algorithm, based solely on the characteristics of data [18]. In this work, we use the latter method; that is, we use feature selection *independent* of any learning algorithm. Hence, we use machine learning and classification accuracy as a *measure* to evaluate feature subsets. The five feature selection methods are discussed in the following.

3.1. V value

The first feature selection (FS) technique we used is the V value method. A material dataset matrix with two classes of materials *class-1* and *class-2* is given. Assume further that we have a threshold value t.For each feature f_i we define four values a, b, c, and $d$as follows:

$a =$ number of materials in *class-1* with values of feature $f_i \geqslant$ t

$b =$ number of materials in *class-1* with values of feature $f_i <$ t

$c =$ number of materials in *class-2* with values of feature $f_i \geqslant$ t

$d =$ number of materials in *class*-*2* with

$$\text{values of feature } f_i < t \tag{1}$$

For a given threshold t, the most useful feature is the one that has the highest absolute difference between $(a+d)$ and $(b+c)$. Then, the measure $|(a+d)-(b+c)|$ is a good indicator of how much a feature differentiates between the two classes. Thus, we compute for each material a V score using our method as follows:

$$V = |(a+d) - (b+c)| \tag{2}$$

This method selects the features that highly discriminates *class-1* from *class-2*. For more details of this method see [4].

3.2. Chi-square (X^2)

The X^2 values are calculated using the above defined a, b, c, d values computed in eq.(1) above [15,53]. Then the formula for X^2 is as follows:

$$X^2 = \frac{n \times (a, d - b, c)}{(a+c) \times (b+d) \times (a+b) \times (c+d)} \tag{3}$$

where n is the total number of materials.

3.3. Information Gain (*IG*)

The information gain IG, for a given feature f_i, and for a given class C, computes the reduction in uncertainty about the value of the class C when the value of the feature f_i is known. According to a, b, c, d values defined in Eq. (1) above, IG can be simplified as follows:

$$\begin{aligned} IG = {} & \frac{a}{n} \log \frac{n \times a}{(a+b) \times (a+c)} \\ & + \frac{b}{n} \log \frac{n \times b}{(b+d) \times (a+b)} \end{aligned} \tag{4}$$

and n is the total number of materials [11,15,53].

3.4. Gain Ratio (*GR*)

From a given set of features, the gain ratio (GR) measure chooses the features that maximize the ratio of gain to its entropy. That is, GR = IG/H(i), where H(i) is the entropy of *fi*. The gain ratio, for our application, is computed as follows:

$$\begin{aligned} GR = {} & \frac{1}{\log(a+b)} * (a * \log \frac{(n*a)}{(a+b)*(a+c)} \\ & + b. \log \frac{(n*b)}{(b+d)*(a+b))} \end{aligned} \tag{5}$$

computed based on the a, b, c, d values defined in Eq. (1) above [36,38].

Table 8
The overall accuracy using N. Bayes classifier for each dataset with 30 features selected according to each FS method

Dataset	With all features	Feature selection (FS):				
		V	X^2	IG	GR	CFS
DS1	80.50	83.98	88.95	88.95	88.95	83.43
DS2	88.33	91.67	92.50	92.50	92.50	90.83
DS3	71.40	70.49	80.33	80.33	80.33	72.13
DS4	71.75	81.77	81.77	81.77	77.90	71.75
DS5	84.53	87.29	87.29	87.29	89.50	84.53

Machine learner: naïve Bayes; 30 features; No. of classes: 2.

3.5. *Correlation-based Feature Selection* (*CFS*)

We used the correlation-based feature selection (CFS) proposed by M. Hall [14]. The CFS criterion measures the ability of a feature, or subset of feature, to predict the associated class using means of measuring correlation between variables.

4. Evaluation and discussion

4.1. *Datasets*

The collected material data were grouped into five datasets as shown in Table 4. The materials are divided into two groups based on the basic nature of each material:

(1) Group-1: Basic element materials (such as *Aluminum Al*, *Barium B*, *Lead Pb*).
(2) Group-2: Compound material (such as *Copper Iron Sulfide*, *FiberNide Nickel Foam*, *96% Silica Glass, Tungsten Silicide*, and *Platinum II Sulfide*).

We calculated the *material range* (g/cm^2) for each material as explained in Sec 2, and values of this property for each dataset are shown in the last column of Table 4.

1. The first and main dataset, *DS1*, consists of 181 materials and 250 properties.
2. The second dataset, DS2, contains only the basic element materials, Group-1, and includes 120 materials and 250 features.
3. The third dataset, DS3, with 61 materials, includes the compound materials, Group-2, only. DS3 includes 250 features.
4. The fourth dataset DS4 contains all material with features representing only component element % (e.g. % Aluminum, % lead, . . . etc); see Table 2.
5. The last dataset, DS5, contains the materials with all properties other than the component element %; examples of these properties are shown in Table 2.

Table 6 shows the distribution of material in the 2-class setting and details of these classes are found in Table 5.

4.2. *Experiments and results*

We conducted the evaluation process using subsets with various numbers of features ranging from 3 features to 30 features and with all features. Initially, we conducted the baseline step on the five datasets with all features with the five learners using the two-class setting and the results are shown in Table 7. Table 7 shows the results of learning-classification accuracy (*LCA*) for 25 combinations (5 datasets $\times$ 5 learners). Notice here that we use the *lca* of the five learners as a measure or indicator of the quality of the included features in each experiment on the radiation shielding. For example, as shown in Table 7, the accuracy with the dataset *DS1*, which includes all materials, and the $J48$ is 83.7% which implies that the quality of this combination of features in correlation with the radiation shielding capacity of the included materials is 83.7%. Thus, $J48$ was able to assign these materials using all features into two classes and was correct 83.7% of the time. Since the accuracy is not perfect, and some materials were assigned the wrong class, based on their feature data, then feature selection mechanism can be used to determine the significance of the various features. In other words, this experiment is meant to examine how well machine learning can distinguish between the two classes of materials where materials in each class have similar radiation shielding capabilities. It is evident that the classifiers can recognize classes and distinguish reasonably between the two classes. In these experiments, the best performer in this experiment is the *Decision Table* with an average 85.2%.

Table 9
The learning-classification accuracy using all classifiers for dataset DS1 with 3, 5, 10, 15, and 30 features selected according to IG method

	All Features	Number of features:				
		3	5	10	15	30
NaiveBayes	80.50	89.27	83.98	85.64	88.40	88.95
L ibSVM	84.70	83.98	87.85	88.87	85.22	87.19
Decision Table	87.50	90.61	92.41	93.04	93.04	91.16
RandomForest	84.16	84.53	85.08	87.29	88.40	91.16
J48	83.71	81.77	88.40	90.61	91.14	91.77

Feature selection: IG.

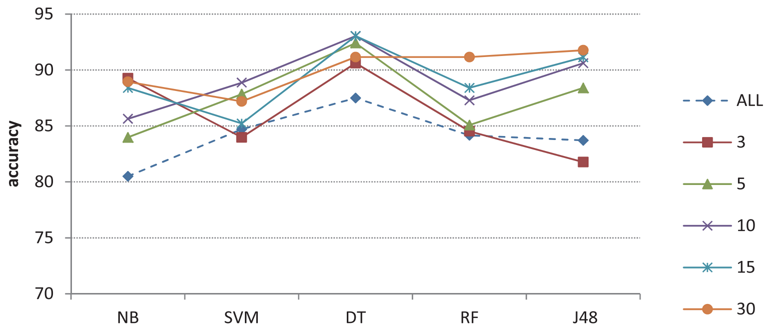

Fig. 1. Performance results of 5 learners using 3 to 30 features selected by IG and dataset DS1.

Table 8 shows the accuracy results of using a simple classifier, *Naïve Bayes* (*NB*), for all datasets and with 30 features selected using various FS methods. These results. This experiment includes 25 combinations (5 FS methods × 5 datasets) plus accuracy with *all-features*. For example, with the combination (X^2, DS1), NB produced 88.95% accuracy ; and similarly, with (*IG, DS1*) and (*GR, DS1*) same accuracy 88.95% compared with the accuracy with DS1 and *all-features* which is 80.50% (Table 8); this indicates that NB can distinguish between the two classes of radiation with 30 features significantly better than with all features.

Table 9 shows the details of using the *IG* method for selecting 3 features, then 5 features, then 10, 15, and 30 features with the five different learners and dataset *DS1*; these results are also illustrated in Fig. 1. We used *IG* here since it produced the best features overall. For example, the last column in Table 9 show that with 30 features, the accuracy results from the five learners are significantly higher than with all features. In general, in this evaluation, there are 25 combinations (learner × no. of features) examined; 23 out of these 25 times; the reduced features produced better performance than all features (see Fig. 1; notice that, in all figures, the dotted line is used for '*all features*'). This implies that all five learners can induce more robust separation models with the reduced features than with all features. More detailed results are also shown in Table 12. These two experiments (Tables 8 and 9) prove that this technique can do fairly well, overall, in reducing features. That is, the feature selection techniques can effectively weed out most of the insignificant features that are not affecting the shielding capabilities of the materials. Table 12 contains the accuracy results of 180 experiments performed on dataset *DS1* in the two-class setting. In this test, *Decision Table* shows the best overall accuracy with most of the feature settings (e.g., with 3 features, with 5 features,..etc).

This set of experiments is repeated for the remaining four datasets *DS2*, *DS3*, *DS4*, and *DS5* (Figs 1, 2, 3 and Tables 7–12). These results represent a large collection of extensive evaluation and testing conducted on five datasets, with five different learners, and with various number of features. These feature subsets were selected using five different techniques and the materials in each dataset are divided into two categories based on their radiation shielding effectiveness. From these result tables (e.g. Table 12) we extracted some significant and useful results and depicted them in Figs 2 and 3. Figure 2 illustrates the accuracy with various combinations of (learner × no. of features) against 'all features' using X^2 for feature selection. In most of the combinations, the reduced features gives better performance than the '*all feature'* case which is represented with the dotted line (Fig. 2). Figure 3, on the other hand, illustrates the results with 15 features against the accuracy

Table 10
Top 40 features with their rankings in all FS methods

Feature Name	V-val	X^2	IG	GR	mean
Density (g/cc)	1	1	1	5	2
Atomic Number	6	2	2	11	5.2
X-ray Absorption Edge	4	3	3	12	5.5
Specific Heat Capacity	11	4	4	6	6.2
Heat of Fusion (J/g)	8	5	8	22	10.8
Melting Point	2	17	17	26	15.5
Electrical Resistivity (ohm-cm)	3	22	24	25	18.5
Poissons Ratio	17	19	20	28	21
Molecular Weight (g/mol)	24	16	19	27	21.5
Electrode Potential (V)	20	24	27	29	25
Thermal Neutron Cross Section	15	29	31	31	26.5
Boiling Point	28	28	30	30	29
Electrochemical Equivalent	29	32	32	32	31.2
Reflection Coefficient, Visible (0-1)	27	84	84	84	69.8
Hardness, Brinell	25	87	87	87	71.5
Hardness, Rockwell A	32	85	85	85	71.8
Critical Magnetic Field Strength, Oersted	30	86	86	86	72
Critical Superconducting Temp (K)	26	89	89	89	73.2
Bulk Modulus (GPa)	31	88	88	88	73.8
Modulus of Elasticity (GPa)	10	96	96	96	74.5
Thermal Conductivity (W/m-K)	5	99	99	99	75.5
Ionic Radius	13	97	97	97	76
Tensile Strength, Ultimate (MPa)	12	98	98	98	76.5
Compressive Yield Strength (MPa)	38	90	90	90	77
Electronegativity	9	100	100	100	77.2
CTE, linear	7	101	101	101	77.5
Hardness, Rockwell B	33	93	93	93	78
Hardness, Rockwell C	37	92	92	92	78.2
Heat of Formation (kJ/mol)	40	91	91	91	78.2
a Lattice Constant	35	94	94	94	79.2
Refractive Index	34	95	95	95	79.8
Elongation at Break (pct)	21	102	102	102	81.8
Emissivity (0-1)	23	103	103	103	83
Hardness, Vickers	18	105	105	105	83.2
Shear Modulus (GPa)	16	106	106	106	83.5
Tensile Strength, Yield (MPa)	22	104	104	104	83.5
Magnetic Susceptibility	14	108	108	108	84.5
Heat of Vaporization (J/g)	19	107	107	107	85
Hardness, Knoop	39	124	124	124	102.7
Critical Temperature	36	126	126	126	103.5

with all features. The results shown in Fig. 3 prove that in all combinations (*learners* × *FS techniques*) the case with 15 features gives better separation than with all features. By and large, the results are fairly good and encouraging. For example, the three features selected using X^2, received accuracy of 92.4% from *DT* learner (Fig. 2 and Table 12) which is impressive as it shows that these three material properties are highly in-line with the radiation shielding range of the tested materials (this same learner, DT, makes 87.5% accuracy with all features). Moreover, the 30 features selected by any FS method (Fig. 2 and Table 12) received always higher accuracy than with all features by all learners.

Experiments with 5 class setting: we conducted brief evaluation for the 5-Class setting. These experiments are done using the same five datasets but with five classes.

Table 11
The learning-classification accuracy results in the 5-class setting using all learners, dataset DS1, and with 3, 5, 10, 15, and 30 features selected according to IG method

Learner	Number of features:				
	3	5	10	15	30
NaiveBayes	80.66	75.69	73.66	73.66	76.14
L ibSVM	75.69	75.69	78.43	78.43	79.01
DecisionTable	80.66	80.66	81.77	81.22	82.87
RandomForest	73.48	80.66	84.81	87.34	80.11
J48	80.11	83.98	83.43	83.98	83.43

Feature selection: IG; dataset: DS1; 5 classes.

Firstly, we ran the five machine learners with specific number of features and the results are in Table 11. The

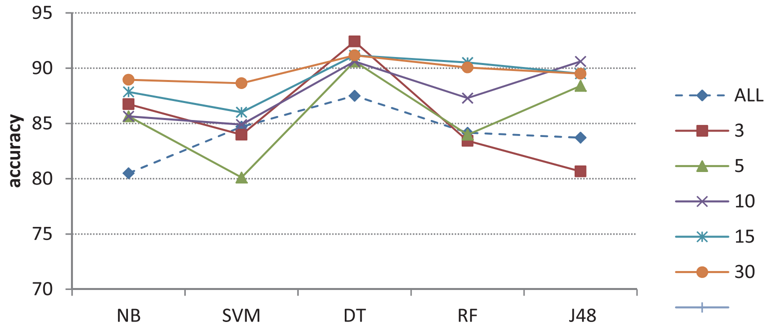

Fig. 2. Accuracy results of 5 machine learners using DS1 with 3 to 30 features and X^2 for feature selection.

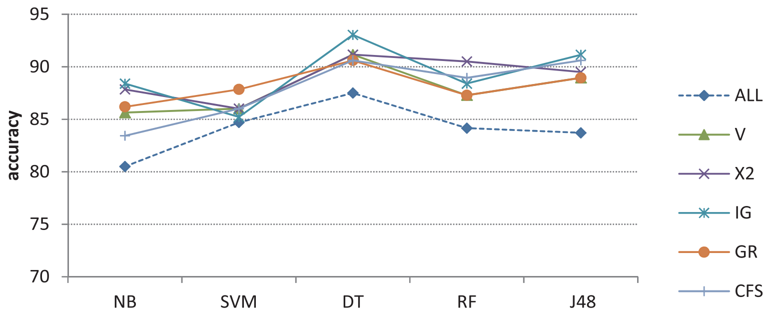

Fig. 3. Results with 15 features selected using all FS methods with the dataset DS1.

Random forest and J48 classifiers performed the best in this set of experiments. Table 11 shows how the *IG* method performs, for all datasets and with every learner. Again, with 30 features, all learners produce higher accuracy results compared with all features as shown in Table 11.

These experimental results and evaluations, with the two classes setting, showed that there is a practically better distinction between the two classes of shielding goodness with small number of features. For example, using the dataset *DS1*, with 10 features selected using *GR* method, we achieved 93% accuracy using DT learner (Table 12) while with all features, DT produce 87.5% accuracy. For example, in Table 12, we notice that, the 3 selected features, achieve with J48 and Random forest about 89% and 90% accuracy respectively with CFS which implies that CFS suggests that the 3 features: *density*, *Atomic Number*, *Heat of Fusion* are the most significant features for radiation shielding property of the tested materials. While, the *density* feature is fairly well known that it affects shielding [7,19,29], the *Heat of Fusion* can be considered a new finding that worth reporting as most of the experiments, with most of the materials, found that *Heat of fusion* is highly correlated with the shielding capability (Table 10).

When the dataset DS1 was divided into five classes of radiation capability, the classifiers, e.g., *Random forest*, were able to achieve 87.3% accuracy for the 15 features, whereas J48 gave 86.2% for the 5 features. When the 120 data elements (*i.e. materials*) in dataset *DS2* were grouped into two classes (good shielding vs. bad shielding), several learners were able to achieve 93.3% to 94.2% for the top 5 selected features. All these results indicated that, machine learning can give higher performance and better accuracy scores for all reduced features subsets selected using five different techniques. Figure 3 also shows that in almost all combination of number of features (3 to 30 features) and learners, always the reduced number of features produce higher accuracy.

Table 10 contains the topmost features selected by four feature selection technique and the mean ranking in the last column. Of course, the ranking of these features are computed based on the equations in Section 3. As we see, all FS methods selected *Density* as the highest ranking features related to radiation shielding. This supports the correctness of the approach as the *density* has been known for long time as the most important property related radiation shielding effectiveness of materials [7,19,29]. Moreover, besides *Density*, other material properties like *Heat of Fusion, Atomic Number*, *X-ray Absorption Edge*, *Electrical Resistivity*, and *Specific Heat Capacity* have been ranked high

Table 12
The learning-classification accuracy (2 classes) using only 3, 5, 10, 15, or 30 features versus all features with dataset DS1 and various classifiers and different FS methods

Dataset DS1; 181 × 250 – Two class setting						
Machine Learner	**Accuracy %**					
	All Features	**With 3 features selected using FS method:**				
		V	X^2	IG	GR	CFS
NaiveBayes	80.50	76.85	86.74	89.27	88.40	83.43
L ibSVM	84.70	77.75	83.98	83.98	83.98	79.01
DecisionTable	87.50	78.27	92.41	90.61	90.61	90.61
RandomForest	84.16	81.01	83.43	84.53	84.53	88.95
J48	83.71	80.14	80.66	81.77	81.77	90.61
	All F	**With 5 features selected using FS method:**				
		V	X^2	IG	GR	CFS
NaiveBayes	80.50	83.98	85.64	83.98	77.35	83.43
L ibSVM	84.70	79.19	80.11	87.85	86.74	79.01
DecisionTable	87.50	90.06	90.61	92.41	87.50	90.61
RandomForest	84.16	85.64	83.98	85.08	82.32	88.95
J48	83.71	83.71	88.40	88.40	80.66	90.61
	All features	**With 10 features selected using FS method:**				
		V	X^2	IG	GR	CFS
NaiveBayes	80.50	82.87	85.64	85.64	81.22	83.43
L ibSVM	84.70	85.54	84.90	88.87	86.74	87.01
DecisionTable	87.50	90.61	90.61	93.04	92.61	90.61
RandomForest	84.16	85.64	87.29	87.29	87.85	88.95
J48	83.71	87.29	90.61	90.61	88.95	90.61
	All F	**With 15 features selected using FS method:**				
		V	X^2	IG	GR	CFS
NaiveBayes	80.50	85.64	87.85	88.40	86.19	83.43
L ibSVM	84.70	86.0191	86.01	85.22	87.85	86.01
DecisionTable	87.50	91.16	91.16	93.04	90.61	90.61
RandomForest	84.16	87.29	90.51	88.40	87.29	88.95
J48	83.71	88.95	89.50	91.14	88.95	90.61
	All F	**With 30 features selected using FS method:**				
		V	X^2	IG	GR	CFS
NaiveBayes	80.50	83.98	88.95	88.95	88.95	83.43
L ibSVM	84.70	87.25	88.64	87.19	87.19	86.01
DecisionTable	87.50	91.16	91.16	91.16	91.16	90.61
RandomForest	84.16	87.29	90.06	91.16	88.95	88.95
J48	83.71	89.50	89.50	91.77	90.06	90.61

and proved computationally to be highly related to the radiation shielding capability of materials.

Clearly, the improved *LCA* accuracy when features are reduced implies that the removed features are noise or redundant and the reduced feature set includes more significant features. As explained in Section 3, if $LCA(DS_i, F^x) > LCA(DS_i, F^y)$ (i.e. the learner produces higher accuracy with feature set F^x than with feature set F^y), then F^x includes more significant features than F^y. Further, if $F^x \subset F^y$ then the extra features in F^y are noise or redundant. The two classes in our application are radiation shielding where C_1 includes the materials with good shielding and C_2 contains materials with bad shielding capability. As shown in Table 12, all five machine learners produce higher accuracy with the reduced feature sets. That is, these learners can induce better separating models with reduced features.

Thus the reduced features have higher correlation with radiation shielding.

We examined the literature to check the validity of the results of our methodology. We researched the material science [2,43] for any physical or material experiments or research work on radiation shielding with any of the most significant features that we found with this methodology. It has been proved in the literature that *atomic number* is property that is correlated with radiation shielding capacity of materials [12]. *Electrical resistivity* is another property that has been confirmed in the literature that is related to radiation shielding [28, 32]. *Electrical resistivity* is inversely proportional with radiation shielding. As the electrical resistivity increases the radiation shielding capability decreases. Also it can be inferred from the study in [12] that *electrical resistivity* is related to radiation shielding as it is proved that intercalation can improve the radiation shielding and significantly lower the electrical resistivity. Similarly, the relation between *Heat of fusion* and shielding is discussed in [45]. Also *Specific heat capacity* (*shc*) of materials has been experimented and reported in the literature and shown to be inversely proportional to radiation shielding and density [24,30,34]. For instance, the *shc* of *lead* ($= 0.129$ kj/kg k) is significantly lower than that of water ($= 4.18$ kj/kg k).To conclude, the top most features that has been identified by most fs techniques are correlated with radiation shielding classes as confirmed by all learners with large number of combinations and also verified from the literature. Some of these features, like *density*, are easy to understand its relatedness with radiation shielding, however, other features can be deemed interesting outcomes of this research.

5. Conclusion

We presented a computational model for identifying the material properties that are mostly related to radiation effectiveness of materials. This work can be considered as a start of a new direction of research in working with radiation shielding using mining and learning based computational techniques. The novelty of this work is in the application of feature selection and learning for understanding the relationships between material properties and radiation shielding. We use a semi-empirical model to compute the radiation shielding effectiveness of each material as a numeric value. We employ the advances in feature selection to identify small subsets of features that are related to the classes of radiation shielding. In other words, these subsets represent the material properties that influence the radiation shielding capability of materials in every class the most. The feature selection and mining techniques are used to analyze material properties *independently* of any learning algorithm; however, we use machine learning and classification accuracy as a measure to evaluate feature subsets. That is, the learners are employed to estimate how well each material can be placed into its correct class of shielding capability based on reduced features. This approach is fairly easy to implement and can be applied in other similar domains. The experimental results proved that we can produce fairly significant results related to this application. These results can be presented as an augmentation and support the discoveries in the similar research in material science and engineering. For example, we have identified some material properties, besides *Density*, like *Heat of Fusion, Atomic Number*, *X-ray Absorption Edge*, *Electrical Resistivity*, and *Specific Heat Capacity* that have been computationally proved to be highly related to, and can improve, the radiation shielding capability of materials. The evaluation results proved that all machine learning algorithms can induce and build more robust separation models for the subsets of reduced number of features that are highly significant in the domain of radiation shielding.

References

[1] E.S. Abouel Nasr and H. Al-Mubaid, Mining Process Control Data Using Machine Learning, CIE, Troyes, France, 2009, pp. 1446–1451.

[2] H. Adeli and S. Kumar, Distributed Genetic Algorithms for Structural Optimization, *Journal of Aerospace Engineering* **8**(3) (1995), 156–163.

[3] H. Al-Mubaid, E.S. Abouel Nasr and M. Hussein, A Methodology for Mining Material Properties with Unsupervised Learning. IJRM, *Int'l J Rapid Manufacturing* **1**(2) (2009), 237–252.

[4] H. Al-Mubaid and N. Ghaffari, A New Gene Selection Technique Using Feature Selection Methodology, Proc. of Int'l Conf on Computers and Their Applications, CATA-2006. Seattle, WA, USA, 2006.

[5] Z. Banković, J.M. Moya, Á. Araujo, D. Fraga, J.C. Vallejo and J.-M. de Goy, Distributed Intrusion Detection System for Wireless Sensor Networks based on a Reputation System coupled with Kernel Self-Organizing Maps, *Integrated Computer-Aided Engineering* **17**(2) (2010), 87–102.

[6] L. Breiman, Random Forests, *Machine Learning* **45**(1) (2001), 5–32.

[7] George Chabot, Health Physics Society, Radiation Basics -Radiation Shielding, http://hps.org/publicinformation/ate/q1094.html.

[8] B. Cyganek, Colour Image Segmentation with Support Vector Machines: Applications to Road Signs Detection, *International Journal of Neural Systems* **18**(4) (2008), 339–345.

[9] M. Dash and H. Liu. Feature Selection for Classification. Intelligent Data Analysis, 1997.

[10] Engineers Edge: http://www.EngineersEdge.com

[11] G. Forman, An Extensive Empirical Study of Feature Selection Metrics for Text Classification, *Journal of Machine Learning Research* **3** (2003), 1289–1305.

[12] J.R. Gaier, J.R. Bunch and M.L. Davidson, Effect of intercalation on the ionizing radiation shielding of graphite fiber composites. NASA Lewis Research Center, Cleveland, USA.

[13] X. Gao, L.P.B. Vuong and M. Zhang, Detecting Data Records in Semi-Structured Web Sites Based on Text Token, *Integrated Computer-Aided Engineering* **15**(4) (2008), 297–311.

[14] M.A. Hall, Correlation-based Feature Selection for Machine Learning. Waikato University, New Zealand, 1999.

[15] B.C. How and N.K. An Empirical Study of Feature Selection for Text Categorization based on Term Weightage, Proc. of IEEE/WIC/ACM Intl Conf on Web Intelligence (WI'04). 2004

[16] P. Huang and Y. Xu, SVM-based learning control of space robotic capturing operation, *International Journal of Neural Systems* **17**(6) (2007), 467–477.

[17] S.L. Hung and H. Adeli, A Parallel Genetic/Neural Network Learning Algorithm for MIMD Shared Memory Machines, *IEEE Trans Neural Networks* **5**(6)(1994), 900–909.

[18] P. Langley, Selection of relevant features in machine learning. Proc of the AAAI Fall Symposium on Relevance, 1994, pp. 1–5.

[19] Lead Industries Association, A guide to the use of lead for radiation shielding, http://www.canadametal.com/pdf/radiation_shielding.pdf.

[20] G. Lebrun, C. Charrier, O. Lezoray and H. Cardot, Tabu Search Model Selection for SVM, *International Journal of Neural Systems* **18**(1) (2008), 19–31.

[21] H. Lee, E. Kim and M. Park (2007), A genetic feature weighting scheme for pattern recognition, *Integrated Computer-Aided Engineering,* 14:2, pp. 161-171.

[22] LET ver 1.24, *Brookhaven National Laboratory*, http://tvdg10.phy.bnl.gov/

[23] Y. Li, Predicting materials properties and behavior using classification and regression trees, *Materials Science and Eng* **433**(1–2) (2006), 261–268.

[24] X. Lia, L.G. Tabila, I.N. Oguochab and S. Panigrahia. Thermal diffusivity, thermal conductivity, and specific heat of flax fiber–HDPE biocomposites at processing temperatures, *Composites Science and Technology* **68**(7–8) (2008), 1753–1758.

[25] H.C. Lian, No-reference Video Quality Measurement with Support Vector Regression, *International Journal of Neural Systems* **19**(6) (2009), 457–464.

[26] H.C. Lian and B.L. Lu, Multi-view gender classification using multi-resolution local binary patterns and support vector machines, *Int'l Journal of Neural Systems* **17**(6) (2007), 479–487.

[27] D.T. Lin and D.C. Pan, Integrating a Mixed-Feature Model and Multiclass Support Vector Machine for Facial Expression Recognition, *Integrated Computer-Aided Engineering* **16**(1) (2009), 61–74.

[28] X. Luo and D.D.L. Chung, Electromagnetic interference shielding using continuous carbon-fiber carbon-matrix and polymer-matrix composites, *Composites: Part B* **30** (1999), 227–231.

[29] M. Mahdy, P.R.S. Speare and A.H. Abdel-Reheem, Shielding Properties of heavyweight, high strength concrete. 2nd Material Specialty Conf of Canadian Society for Civil Eng, June 2002.

[30] Material properties at Engineering toolbox: www.engineering toolbox.com.

[31] Metal Science Resource, http://www.lib.iastate.edu/collecti ons/eresourc/matsci.html.

[32] Nasa Glenn Research Center. Radiation Protection of New Lightweight Electromagnetic Interference Shielding Materials Determined. http://www.grc.nasa.gov/WWW/RT/RT1995/5000/5480g.htm.

[33] NASA Space Radiation Laboratory at Brookhaven: http://www.bnl.gov/medical/NASA/NSRL_description.asp.

[34] Physics Forums: www.physicsforums.com.

[35] A. Ponomarev, H. Nounu, H. Hussein, M.-H. Kim, W. Atwell and F. Cucinotta, NASA-developed ProE-based tool for the ray tracing of spacecraft geometry to determine radiation doses and particle fluxes in habitable areas of spacecraft and in the human body. NASA/TP-2007-214770, 2007.

[36] J.R, Quinlan, Induction of Decision Trees, *Machine Learning* **1** (1986), 81–106.

[37] J.R. Quinlan, C4.5: Programs for machine learning. Morgan Kaufmann, San Mateo, CA, USA, 1993.

[38] J.R. Quinlan, Improved Use of Continuous Attributes in C4.5, *Journal of Artificial Intelligence Research* **4** (1996), 77–90.

[39] R.T.K. Lin, J.L.T. Chiu, H.-J. Dai, R.T.-H. Tsai, M.-Y. Day and W.-L. Hsu, A Supervised Learning Approach to Biological Question Answering, *Integrated Computer-Aided Engineering* **16**(3) (2009), 271–281.

[40] M. Salick. Radiation Protection from Solar Particle Events (SPE's), Galactic Cosmic Rays (GCR's), and Other Solar Activity for Space Missions. UW-Madison, Engineering Mechanics and Astronautics, February 2006.

[41] K. Sarma and H. Adeli, Fuzzy Discrete Multicriteria Cost Optimization of Steel Structures, *Journal of Structural Engineering, ASCE* **126**(11) (2000), 1339–1347.

[42] K.C. Sarma and H. Adeli, Life-Cycle Cost Optimization of Steel Structures, *Int'l Journal for Numerical Methods in Engineering* **55**(12) (2002), 1451–1462.

[43] K. Sarma and H. Adeli, Fuzzy Genetic Algorithm for Optimization of Steel Structures, *Journal of Structural Engineering* **126**(5) (2000), 596–604.

[44] C. Silva and B. Ribeiro, Towards Expanding Relevance Vector Machine to Large Scale Datasets, *International Journal of Neural Systems* **18**(1) (2008), 45–58.

[45] L.E. Steiner and J. Johnston, Development of a method of radiation lamorimetry and the heat of fusion or of trasition of certain substances, *J Phys Chem* **32**(6) (1928), 912–939.

[46] The material web: http://www.matweb.com.

[47] The SAO/NASA Astrophysics Data System: http://adsabs.harvard.edu/.

[48] M. Vasta, R. Singh and A. Noore, Integrating Image Quality in 2ν-SVM Biometric Match Score Fusion, *International Journal of Neural Systems* **17**(5) (2007), 343–351.

[49] E.D. Wandekokem, E. Mendel, F. Fabris, M. Valentim, R.J. Batista, F.M. Varejao and T.W. Rauber, Diagnosing multiple faults in oil rig motor pumps using support vector machine classifier ensembles, *Integrated Computer-Aided Engineering* **18**(1) (2011), 61–74.

[50] I.H. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes and S.J. Cunningham, Weka: Practical Machine Learning Tools and Techniques with Java Implementations.

[51] Y. Yang and B.L. Lu, Protein subcellular multi-localization prediction using a min-max modular support vector machine, *Int'l Journal of Neural Systems* **20**(1) (2010), 13–28.

[52] V. Zajic and P. Thieberger, Heavy Ion Linear Energy Transfer Measurements during Single Event Upset Testing of Electronic Devices, *IEEE Transactions on Nuclear Science* **46** (1999).

[53] Z. Zheng and R. Srihari, Optimally Combining. Positive and Negative Features for Text Categorization. ICML, 2003.

[54] J.F. Ziegler, Stopping Cross-sections for Energetic Ions in All Elements, Vol. 5 of The Stopping Powers and Ranges of Ions in Matter, Pergamon Press, New York, 1977.