Artificial Intelligence in Plasma Cell Myeloma: Neural Networks and Support Vector Machines in the Classification of Plasma Cell Myeloma Data at Diagnosis

Ashwini K. Yenamandra, Caitlin Hughes, Alexander S. Maris

Department Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA

Abstract

Background: Plasma cell neoplasm and/or plasma cell myeloma (PCM) is a mature B-cell lymphoproliferative neoplasm of plasma cells that secrete a single homogeneous immunoglobulin called paraprotein or M-protein. Plasma cells accumulate in the bone marrow (BM) leading to bone destruction and BM failure. Diagnosis of PCM is based on clinical, radiologic, and pathological characteristics. The percent of plasma cells by manual differential (bone marrow morphology), the white blood cell (WBC) count, cytogenetics, fluorescence *in situ* hybridization (FISH), microarray, and next-generation sequencing of BM are used in the risk stratification of newly diagnosed PCM patients. The genetics of PCM is highly complex and heterogeneous with several genetic subtypes that have different clinical outcomes. National Comprehensive Cancer Network guidelines recommend targeted FISH analysis of plasma cells with specific DNA probes to detect genetic abnormalities for the staging of PCM (4.2021). Recognition of risk categories through training software for classification of high-risk PCM and a novel way of addressing the current approaches through bioinformatics will be a significant step toward automation of PCM analysis. **Methods:** A new artificial neural network (ANN) classification model was developed and tested in Python programming language with a first data set of 301 cases and a second data set of 176 cases for a total of 477 cases of PCM at diagnosis. Classification model was also developed with support vector machines (SVM) algorithm in R studio and interactive data visuals using Tableau. **Results:** The resulting ANN algorithm had 94% accuracy for the first and second data sets with a classification summary of precision (PPV): 0.97, recall (sensitivity): 0.76, f1 score: 0.83, and accuracy of logistic regression of 1.0. SVM of plasma cells versus TP53 revealed a 95% accuracy level. **Conclusion:** A novel classification model based only on specific morphological and genetic variables was developed using a machine learning algorithm, the ANN. ANN identified an association of WBC and BM plasma cell percentage with two of the high-risk genetic categories in the diagnostic cases of PCM. With further training and testing of additional data sets that include morphologic and additional genetic rearrangements, the newly developed ANN model has the potential to develop an accurate classification of high-risk categories of PCM.

Keywords: Artificial neural network, cytogenetics, fluorescence *in situ* hybridization, machine learning, microarray, National Comprehensive Cancer Network, Next Generation Sequencing, plasma cell myeloma, support vector machines kernel trick

INTRODUCTION

Plasma cell neoplasm and/or plasma cell myeloma (PCM) is a group of mature B-cell disorders characterized by a clonal expansion of plasma cells (a type of white blood cell [WBC] called a plasma-B cell) that secretes a single homogeneous immunoglobulin called paraprotein or M-protein.[1] Clinical diagnostic criteria include hypercalcemia, renal insufficiency, anemia, and bone lesions (CRAB criteria).[2]

Access this article online

Quick Response Code:

Website:
www.jpathinformatics.org

DOI:
10.4103/jpi.jpi_26_21

PCM accounts for 1.8% of all cancers and 17% of hematological malignancies in the United States and is most frequently diagnosed among adults aged 65–74 with a median age of 69 years.[3] The American Cancer Society estimated 32,270 new cases (17,530 in men and 14,740 in women) and close to 12,830 deaths (7,190 in men and 5,640 in women) in 2020.[4] The median survival rate is usually 5 years with only 10% of patients that live 10+ years.[5,6] Survival rates depend mostly on the level of serum-2-macroglobulin, albumin, M-protein, calcium, creatinine, and presence or absence of bone lesions.[3]

PCM is a complex disease with different clinical phases and various risk levels.[2-10] Clinical phases consist of monoclonal gammopathy of undetermined significance (MGUS), smoldering multiple myeloma and plasma cell.[2-10] Plasma cells accumulate in bone marrow (BM) leading to bone destruction and marrow failure.[10-16] Diagnosis is based on clinical, radiologic, and pathological characteristics.[7-8] For risk stratification, prognosis, and treatment efficacy, PCM is classified into high, standard, and low-risk clinical categories based on serum M-protein concentration, percent of plasma cells in the marrow (extent of bone marrow involvement), and identification of genetic abnormalities.[9-16] The genetic abnormalities usually reflect various underlying pathways of clonal heterogeneity and subsequent evolution.[8-11] The genetic alterations are critical for prognosis, risk stratification, expected patient outcome, survival rate, and in selecting an appropriate therapeutic strategy.[9] High-risk cytogenetics and persistent minimal residual disease (MRD) by flow cytometry may predict relapse after autologous stem cell transplant (ASCT; National Comprehensive Cancer Network [NCCN] 4.2021). DNA sequencing, microarray, cytogenetics, and fluorescence *in situ* hybridization (FISH) studies on BM require validation before clinical use (NCCN).[3] Recommendations of NCCN include metaphase cytogenetic analysis of bone marrow, as well as FISH probes on plasma cells to detect abnormalities of del(1p32), 1q21 amplification, del(13q), t(4;14), t(11;14), t(14;16), t(14;20), and del(17p) at the time of diagnosis [Table 1]. High-risk abnormalities include multiple mutations in different pathways, including missense, nonsense, splice-site mutations and deletion of TP53 locus (in this paper referred to as 17p and/or TP53 or P53), t(4;14), t(14;16), t (14;20), and hypodiploidy. Deletion of chromosome

1p, gain of copy number or amplification of chromosome 1q, and abnormalities of TP53 locus have been reported to indicate PCM disease progression.[12-17] Genetic aberrations observed in standard to low-risk PCM include t(11;14), t(6;14), deletion or loss of chromosome 13, and hyperdiploidy.[10,17-20] Deleted 13q is a negative prognostic indicator when observed in metaphase (dividing cells) cytogenetic analysis. Secondary genetic events consist of additional numerical and structural chromosomal abnormalities.[17] Exome sequencing study of more than 1000 samples revealed heterogeneity (20%) and frequent mutations (25%) involving KRAS and NRAS genes.[10] The gene expression profile (GEP) signature is an emerging technology with 16-, 70-, and 92-gene panel models for interrogation of molecular aberrations in PCM; however, GEP is still not currently available in clinical practice for diagnostic workup.[3] Currently, FISH analysis is being used in clinical diagnostic laboratories to identify genetic rearrangements.[10,17-20]

Related work

The outcome and median survival rate of PCM patients has significantly improved by minimizing cytotoxic chemotherapies through the use of autologous stem cell transplantation (ASCT), immunomodulatory therapy (thalidomide, stem cell transplantation (ASCT) and immunomodulatory therapy proteasome inhibitors-bortezomib, ixazomib, and carfilzomib), monoclonal antibody therapy (elotuzumab, daratumumab) and molecularly targeted histone deacetylase inhibitors (HSP90 inhibitors, AKTinhibitors, and KSP inhibitor).[21] However, high-risk patients have shorter progression-free survival.[21] Therefore, earlier identification of genetically high-risk patients to initiate therapy with modern therapeutic agents is one of the important factors in the overall survival.

Identification of chromosomal abnormalities through metaphase analysis has limitations due to the low proliferation of plasma cells in bone marrow cultures. Genetic abnormality is detectable in only 10%–30% of cases due to the low percent of plasma cells and/or low proliferation rate of plasma cells.[12] Currently, targeted FISH is a recommended diagnostic modality for the identification of genetic abnormalities and risk stratification at the time of diagnosis.[3] An increase in the abnormal detection rate was

Table. 1 Correlation of various genetic abnormalities with high and standard risk status in PCM

Gene Loci/ Ploidy	Chromosome band	Type of abnormality	Mutations	Risk status
TP53	17p13	Deletion	Missense, Nonsense and Splice-site	High
IGH/FGFR3	t (4;14)(p13;q32)	Fusion genes	N/A	High
IGH/MAFA	t (14;16)	Fusion genes	N/A	High
IGH/MAFB	t (14;20)	Fusion genes	N/A	High
CKS1B	1q21	Amplification	N/A	High
CDKN2C	1p32	Deletion	N/A	High
Monosomy?	Loss resulting in monosomy of several chromosomes	Hypodiploid	N/A	High
CCND1/IGH	t (11;14)(q13;q32)	Fusion genes	N/A	Standard
CYCLIND2/IGH	t (6;14)(p21;q32)	Fusion genes	N/A	Standard
RB1	Chromosome 13	Deletion/monosomy	N/A	Standard
CN LOH?	Gain of odd-numbered chromosomes is distinct	Hyperdiploid	N/A	Standard

reported by many laboratories through purification or enrichment of plasma cells from BM specimens by the use of RoboSep™ from Stemcell Technologies, Canada [Figure 1]. RoboSep™ is an automated cell processing and enrichment method that involves the use of anti-CD138-coated magnetic beads (immunomagnetic bead technology) to enrich CD138+ plasma cells in patient samples followed by targeted FISH.[20]

In our laboratory, an increase from 30% to 80% was observed in the detection rate of abnormal cases by enrichment of bone marrow plasma cells and subsequent FISH analysis [Figures 2 and 3]. However, the CD138 enrichment procedure is technically challenging as there is no reimbursement (CPT code) for the enrichment procedure for laboratories, can add additional overhead cost to institutions for processing of the specimens for enrichment, and maintenance of reagents. BM specimens obtained from patients need to be processed for critical diagnostic hematopathology procedures including morphology and flow cytometry. As such, bone marrow

specimens received in the cytogenetics/molecular laboratory can be limited for CD138 enrichment and downstream FISH, chromosome analysis, chromosomal microarray (CMA), and/or GEP. Cell pellets obtained from CD138 enrichment are usually small with poor cellular morphology and/or may have weak hybridization signals with targeted FISH probes; thus, FISH needs to be repeated for those probes. Some laboratories may not be able to report a particular FISH probe result due to depletion of the enriched pellet. In our laboratory, we have to choose between high- and low-risk gene rearrangement detection probes or to count a lower number of cells than ideal if the cell pellets are small (in a diagnostic setting, 200 interphase cells are required for each probe) and not report or bill the results. Therefore, recognition of risk categories through training software for classification of high-risk stratification will be a significant step toward automation of PCM analysis for prognostic and therapeutic decisions.

Problem space and motivation

With the emergence and growth of personalized medicine, artificial intelligence (AI) has become an important technology to bring new opportunities in the practice of medicine.[22] Machine learning (ML) and/or deep learning, a subfield of AI, is increasingly used for diagnosis and prediction of diseases such as cancer, diabetes, neurological disorders, and cardiovascular diseases through interrogation of genomic data and turning the data into actionable insights. Current literature on machine learning in PCM is based mostly on clinical trials for predicting treatment benefit, multilevel drug response, ICD-9-CM diagnosis codes, administrative data using Surveillance, Epidemiology, and End Results-NCI (SEER) registry, and distinguishing smoldering versus symptomatic multiple myeloma.[22-29] The application of AI techniques in PCM diagnosis is still at an early and unexplored stage. The purpose of this paper is to explore an ANN algorithm for classification of the PCM data at initial diagnosis into normal
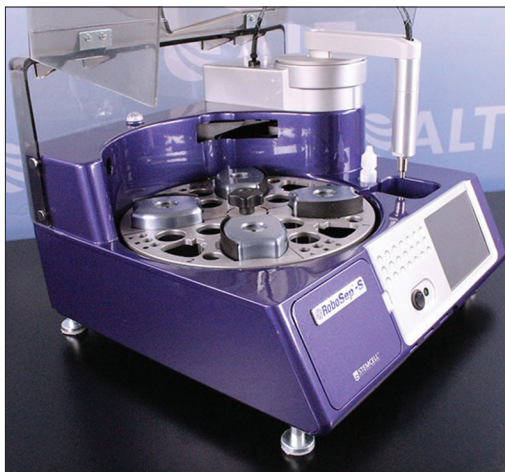

Figure 1: RoboSep™ stem cell technologies Canada

Figure 2: Tableau – Cohort 1 – Abnormal versus normal cases based on age in males and females

and abnormal categories based on results of morphological, flow, and genetic variables of certain high risk alterations included in the study.

Research question

An artificial neural network (ANN) is a mathematical model that is inspired by the way the biological nervous systems, such as the human brain, processes information. ANN's ability to learn quickly is what makes it a powerful and useful tool for a variety of tasks such as classification, pattern recognition, and modeling.[28] This paper aims to introduce bioinformatics, especially ANN to PCM researchers. In the current study, we explored the PCM data to find if hidden patterns in hematological and genetic variables at diagnosis can provide a suitable input for ANN to classify normal and abnormal results. Our research question for this study was:

1. Can a subset of PCM data including age, WBC, and percentage of plasma cells at initial diagnosis through the morphological study of bone marrow, in addition to known high-risk cytogenetic alterations, be used to design a predictive algorithm for identification of risk status?
2. If so, could a classification model aid the critical variables in the disease management initiatives?
3. Would such an approach represent a promising tool for diagnosis and/or follow-up of PCM patients?

Herein, we describe a novel tool using ANN, a deep learning model for a prognostic classification of PCM patients. The data described in this paper was collected only after Institutional Review Board (IRB) approval, was de-identified, and was generated based on the tested hematological and genetic variables at our academic medical center.

Materials and Methods

Sample collection and preparation

Bone marrow samples of the possible PCM cases were received in the hematopathology laboratory at our academic medical institution. Samples were processed for flow cytometry and BM morphology according to standard hematopathology laboratory procedures and the diagnostic results were reported by pathologists. Concurrent bone marrow samples received in the cytogenetics laboratory were processed for cytogenetic analysis and targeted FISH PCM panel following CD138 enrichment with magnetic bead separation technique [Figure 1]. FISH technique involves hybridization of a complimentary DNA sequence probe with a fluorescent tag to the region of interest. The resulting hybrid FISH signal can be visualized under a fluorescence microscope. The FISH probe (Abbott Molecular, Downers Grove, IL) panel set used for the testing consisted of t(4:14), t(11;14), RB1/LAMP1 (13q14/13q34), and TP53/centromere 17 [17p13/17 centromere; Figure 3] in our cytogenetics laboratory. FISH slides were processed and analyzed according to standard cytogenetics laboratory procedures. Two hundred cells per probe were scored by two technologists under the Nikon Fluorescence microscope and micrographs were taken under x100 magnification. Results of all probes were tabulated and were reported by pathologists.
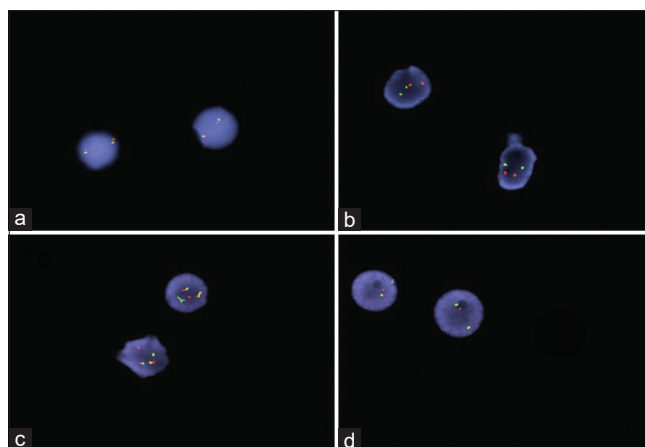

Figure 3: Fluorescence *in situ* hybridization with myeloma-specific probes (a) 13q14 (red)/13q34 (green) probes with normal pattern (b) CCND1 (red)/IGH (green) probes with normal pattern. (c) FGFR3 (red)/IGH (green) probes with rearranged (fusion of red and green) pattern. (d) TP53 (red)/centromere17 (green) with deletion (loss of on red) of TP53 locus

The data was collected from samples at initial diagnosis, of which a small number had a TP53 deletion. Mutations of TP53 were not tested for.

PCM cases included in this study were diagnosed at our academic medical institution. WBC count was determined by complete blood cell count at the time of bone marrow biopsy. Plasma cell percentage was based on manual cell count (bone marrow morphology of at least 200 cells). All the data, including FISH, were de-identified and collected retrospectively only after approval by the IRB. The data consisted of two cohorts, collected at two different time points. The first cohort (cohort 1) consisted of 301 cases collected from July 20, 2017, to August 31, 2018 (IRB # 182023). The second cohort (cohort 2) consisted of 176 cases collected from September 1, 2018, to May 31, 2019 (IRB # 191662). The variables included in the dataset were age, gender, percent of BM plasma cells, WBC, and results of PCM FISH analysis.

Data collection and preprocessing for analysis

Both Cohort 1 and Cohort 2 datasets collected at diagnosis had relatively fewer number of high-risk category variables compared to low- or standard-risk categories [Figure 4]. Another difference between the two datasets is the total number of cases in each dataset. Both the datasets had similar types of variables (results of morphology, flow, tested fish probes, and demographics) except that Cohort 1 was qualitative with a binary numerical value of 0 or 1 for results of various variables and the risk status. The risk status was represented as 1 for normal and 2 for abnormal. Cohort 2 data was quantitative with a percentage of abnormal cells for different variables and binary numerical value for risk status representing 0 for normal and 1 for abnormal.

Cohort 1

- Step 1: Cohort 1 data set consisted of 20 columns with variables corresponding to the result (normal versus abnormal) of each FISH probe
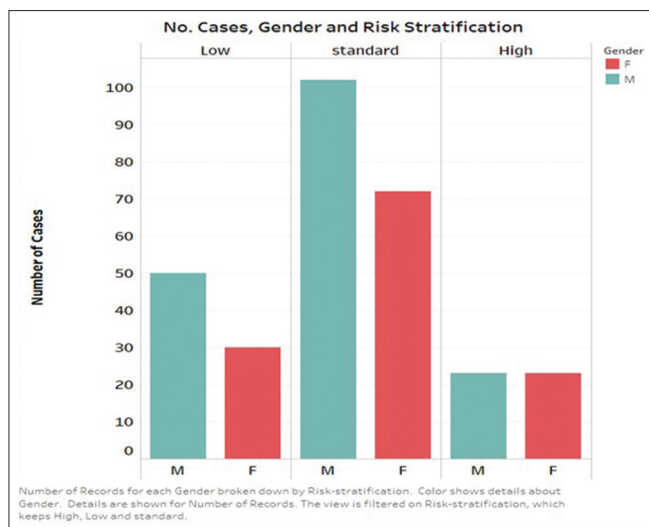
Figure 4: Tableau visual of Cohort 1 data: Male and female cases with low, standard, and high risk, low and standard risk: males higher than females, high risk: males and females similar in number

- Step 2: Data required data preprocessing steps including conversion of the gender (categorical) column into a numerical value (male = 1, female = 2)
- Step 3: Scores for each probe and case result were converted into numerical values of 1 or 2 (normal = 1, abnormal = 2)
- Step 4: The Excel file was converted to a Comma Separated Value (CSV) file to be imported into Python and R studio.

The columns with 20 variables such as case number, gender, age, WBC, percent of BM plasma cells
- Step 5: In addition to the original data set with 20 variables, subsets were also created with a combination of variables
- Step 6: The original data set contained all variables; subsets contained mostly high-risk variables
- Step 7: CSV file containing a subset of cohort 1 data was also imported into Tableau for data visualization and into R studio for SVM with Kernel trick.

Cohort 2

Cohort 2 data set was similar to cohort 1 except that the normal result for a specific FISH probe was left empty, whereas the abnormal result was represented as a numerical percentage of abnormal cells in the respective columns for the variables. The last result column was converted into numerical values of 0 or 1 (normal = 0, abnormal = 1).

Preprocessing and steps involved in the analysis of cohort 2 data were performed as in cohort 1.

RESULTS AND DISCUSSION

Our research questions again were:
1. Can a subset of the PCM data including age, WBC, and percentage of plasma cells at diagnosis, in addition to known high-risk cytogenetic alterations, be used to design a predictive algorithm for identification of PCM risk status?
2. If so, could a classification model based on the critical variables aid in disease management initiatives?
3. Will a comprehensive collection of PCM genomic rearrangements data that impact the progression and treatment efficacy of PCM help in understanding the complexity of PCM and ultimately improve the survival of patients?

To answer the above questions, we designed a predictive algorithm for the identification of risk status by collecting a total of 301 and 176 cases with 20 variables in cohorts 1 and 2, respectively. Subsets were also created from cohorts 1 and 2 for deep learning analysis through Python, R studio, and Tableau to design the models.[30,31] In this paper, we discuss the results of Python-designed ANN models and Tableau analysis.

Selection of network architecture, training/evaluation, and hyperparameters

ANN or NN are a family of computational algorithms with a trainable subfamily of models that can be optimized for several different functions.[24,29,30] Advantages of NN include their high tolerance to noisy linear and nonlinear data and ability to learn quickly and classify patterns on which they have not been trained.[24,29,30] The main advantage of ANN is that it is not based on assumptions, it allows detection of connections between factors in a wide range of problems, and it has given superior results to conventional statistical models in many instances.[24,29,30]

Our intention was to generate NNs and build a predictor for risk status that can be used with test data at time of initial PCM diagnosis. We divided data into two datasets of several combinations: (1) a learning set to build the models and (2) a testing set for the evaluation. For our data which involve risk classification with many variables, we choose NN. Briefly, NN or ANN consists of interconnected layers of algorithms called neurons (nodes or perceptron that feed data into each other) with an activation function, one or more weighted input connections, and a transfer function that combines the inputs and an output connection. The input for each layer is the output of the previous layer. NN is a feedforward network that takes the data points (variables) and classifies the data points. When the model is learning or is being trained, patterns of information from the dataset get fed into the network through the input neurons. This triggers the layers of hidden neurons, and these, in turn, activate the output neuron, with the output of the preceding layer being the input of the subsequent layer. This process of feedforward network is a popular algorithm of ANN. Every neuron adds up all the inputs it receives. If the sum is more than a certain threshold value, the neuron fires and triggers neurons that are connected to its right side. NNs learn exactly the same way the neuronal structure of the mammalian cerebral cortex learns but on a much smaller scale by a feedback process called back-propagation, the next most

popular ANN algorithm. In backpropagation, the algorithm processes data backward from the output through the hidden neurons to the input neurons (goes backward) and causes the network to learn through the difference (also called as error rate or cost function) that is between the predicted and actual outputs. The network then modifies weights of the connections between the neurons and tries to learn the correct output (the category a specific data point belongs to) for classification.

Building the artificial neural network

We built a three-layer network with an input layer corresponding to the selected variables for generating various models. The first hidden layer varied from 4 to 19 nodes depending on the number of input variables and an activation function ReLU and the second hidden layer with 4–8 nodes and ReLU activation function. The number of neurons of the hidden layer was also determined according to the number of variables. The output layer had one node with a sigmoid activation function and one linear output unit to ensure that the network output was between 0 and 1, a binary risk outcome. Weights and biases of NNs were determined by training with a two-phase procedure. The first phase of backpropagation had a moderate training rate and the second phase was a gradient descent, another powerful algorithm of NN. To interpret the network output as probabilities and to make them comparable to the results of logistic regression, we used a cross-entropy error function to adjust weights. Optimization algorithm Adam was used for stochastic gradient descent to train the classifier (model). The model was trained with fit function and trained over 100–1000 epochs with each epoch split into batches. Finally, the model was evaluated on the training dataset with evaluate function to generate a prediction for each input and output accuracy.

ANN was developed in Python programming language using TensorFlow 2 (tensorflow.org) backend and the Keras library (keras.io). The cohort data were divided into "training" and "test," for training and testing the NN model, respectively.

Stepwise procedure of artificial neural network model design of cohort 1 and cohort 2 data

Since the model was to be used for prediction, a simple and stepwise variable selection procedure was implemented in designing various models. The stepwise procedure included creating subsets with the addition or removal of independent variables that were noisy and/or reduced the accuracy for building the model while keeping the variables that fit the data best. In other words, the quality criterion considered was the size of the model (small models that fit well) while eliminating the correlation between predictor variables that can have undesirable effects on models. We built close to 200 models in Python and R studio using the entire dataset of 20 variables and a combination of subset of a number of variables of cohort 1 data. The subsets' names were created randomly. For example, MM1, MM2, MM3, MM4, and MM411 were all combinations of different variables of cohort 1 dataset. Cohort 1 – NN models were built with training and testing data in various combinations of ratios consisting of 90/10,

80/20, 75/25, 70/30, 60/40, and 40/60, respectively. The models including size (number of nodes in the model, width (number of nodes in a specific layer), depth (number of layers in a NN), and various activation functions such as ReLU, Leaky ReLU, and Tanh resulted in slightly different accuracy levels. However, discussion of every model designed is beyond the scope of this paper due to the breadth and complexity. Therefore, only a few models that led to the high accuracy are discussed here. The data set cohort 1 was labeled as MM1 and subsets were labeled as MM10, MM2, MM31, MM411 (cohort 1), and PCMQ3 (subset of cohort 2) for ANN. A brief review of a few of the failed and successful models was discussed below.

Our findings
Cohort 1 and subsets

- MM1 data – Cohort 1 data set with 20 columns, variables were case number, gender, age, WBC, BM percent of plasma cells. The algorithm did not work with or without Tensorflow 2 after the compile ANN step even with different activation functions or with different training and test percentage ratios

- MM10 data – 10 columns, gender, age, WBC, or BM plasma cell percentage were excluded. FISH variables that were considered noisy based on the MM1 data processing (e.g., multiple copies of specific genetic loci) were also excluded. ANN accuracy was 27% and accuracy did not change with either different activation functions or different training or test percentage ratio. TensorFlow 2 did not work either

- MM2 data – 13 columns, age, WBC, and percentage of BM plasma cells were included. Of note, gender at this point was causing several failed models and was excluded at this point. ANN accuracy was 45% and accuracy remained the same with different activation functions and/or different training or test percentage ratio

- MM31 – 6 columns, age, WBC, BM plasma cell percentage, TP53, and FGFR3/IGH. Gender was excluded. ANN accuracy reached 90% with a ratio of 60/40; however, predictions had several false negatives [not included in Figure 5]

- MM411 – At this point, the authors recognized that age and gender were noisy in building the algorithm. Therefore, in the next set MM411, only five columns with WBC, BM plasma cell percentage, TP53 deletion, FGFR3/IGH, and results were used. When the data was divided into 60/40 percent training and test sets, respectively, ANN accuracy reached 99%, but several false positives and FN were classified as true positives (TP) or true negatives.

In the next step, ANN with train/test ratio of 40/60 ratio was built with a 94% accuracy and was effective in building the ANN model [Figures 5 and 6].

Cohort 2 and subsets

PCMQ3: Five columns (the same as in MM411), with WBC, BM plasma cell percentage, TP53 deletion, and FGFR3/IGH. The total number of cases was 176. ANN model that worked

Number	Dataset	Columns	Variables	Training %	Testing%	Accuracy	Cross validation using KNN	ROC curve with Logistic regression
1	Cohort 1 MM10 301 cases	10	Gender, age, WBC and % plasma cells excluded	80%	20%	27%	no	no
2	Cohort 1 MM2- 301 cases	13	Gender excluded, age, WBC, % Plasma cells included	80%	20%	45%	no	no
3	Cohort 1 MM411 301 cases	5	Gender, age excluded, WBC, % Plasma cell included	60%	40%	99% (model predicted mostly the value of the majority class –Accuracy paradox and overfitting)	no	no
4	Cohort 1 MM411 301 cases	5	Gender, age excluded, WBC, % Plasma cell included	40%	60%	100%	94%	1.0

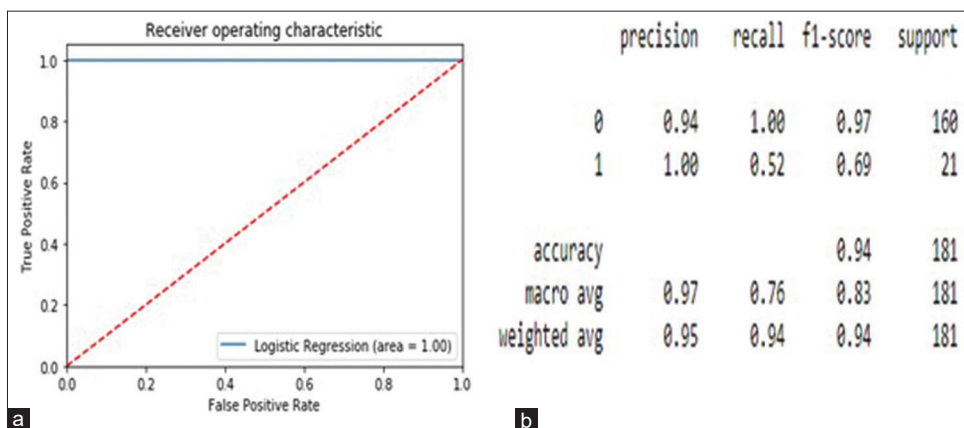**Figure 5:** Python predictive models summary table

Figure 6: PCMQ3 with TensorFlow 2 and ReLU with a ratio of 40/60 training and testing respectively. (a) Receiver operating characteristic curve in logistic regression to determine the best cutoff value for predicting whether a new observation is 0 or 1. (b) Classification report

with MM411 was secondarily validated with the same train/test ratio of 40/60 percent, respectively, and 94% accuracy was obtained for PCMQ3. Of note, 40/60 percent train/test algorithm seems to be the most effective algorithm for this type of study [Figures 5 and 6].

Evaluation of artificial neural network model

Model using MM411 dataset with 60/40 ratio was able to predict only the value of the majority of the class variable for classification. This accuracy problem (also known as accuracy paradox) was probably due to the fact that our data had a slight imbalance for variable distribution (total number of cases that had an abnormal result for high risk was lower than the cases with a normal result). Of note, imbalanced data distribution was not due to inaccurate collection of data but rather due to the fact that the two high-risk categories we have taken into consideration in designing the algorithm are infrequent in PCM cases. This could be attributed to selection bias. Furthermore, the imbalance is probably expected in real-time data with high-risk variables in the population. Additional studies fitting the criteria of high-risk categories that contribute to a better and

final weighted average will certainly help clarify this aspect in future.

Classification summary

For MM411 and PCMQ3 datasets, the classification summary with a training and testing ratio of 40/60 revealed the following values: precision (PPV): 0.97, recall (sensitivity): 0.76, and f1 score: 0.83, and accuracy of logistic regression: 1.0. The 40/60 ratio model seems to be a better model with better predictive ability [Figure 5].

Tableau visualizations

Tableau (2020.3) is an interactive visualization platform to create analytics models using data sets. Visualization with Tableau analytics of cohort 1 data set for the high-risk categories precisely unveiled valuable insights into how and why the ANN worked with the data set.

A comparison of the enriched and unenriched cases indicated that the number of abnormal cases increased with CD138 enrichment; this was true for both males and females [Figure 2]. More cases with increased plasma cells were observed in males

between age 65 and 70, whereas in females, it was between 75 and 80 years of age. Cohort 1 data (all enriched samples) indicated that low- and standard-risk occurred more frequently in males, while in the high-risk category, the number of males and females was similar [Figure 4]. In addition, we observed that high-risk cases with t(4;14) and TP53 loss were relatively associated with lower BM plasma cell percentages and WBC [Figure 7]; forecast indicator (projection) revealed the same [Figure 8]. However, additional studies are necessary to confirm this finding.

Support vector machines

SVM is a labeled training data (supervised learning) algorithm. The objective of SVM is to find a decision boundary (hyperplane) in N-dimensional space (where N is the number of features or variables) and distinctly classify data points. MM4 data set with plasma cell percentage and

TP53 status was used to create SVM Kernel trick in R studio using both 75/25 and 60/40 ratio split. The SVM accuracy was 95% [Figures 9-12].

Addressing the research questions and future work

1. Can a subset of PCM data including age, WBC, and percentage of BM plasma cells at diagnosis through morphological study of bone marrow, in addition to known high-risk cytogenetic alterations, be used to design a predictive algorithm for identification of risk status?

Response: We analyzed and built close to 200 models in Python and R studio using the entire dataset of 20 variables and a combination of subset of variables of the of PCM data including age, WBC, and percentage of BM plasma cells at initial diagnosis, in addition to known high-risk cytogenetic alterations to design a predictive algorithm for identification of risk status at the time of diagnosis.
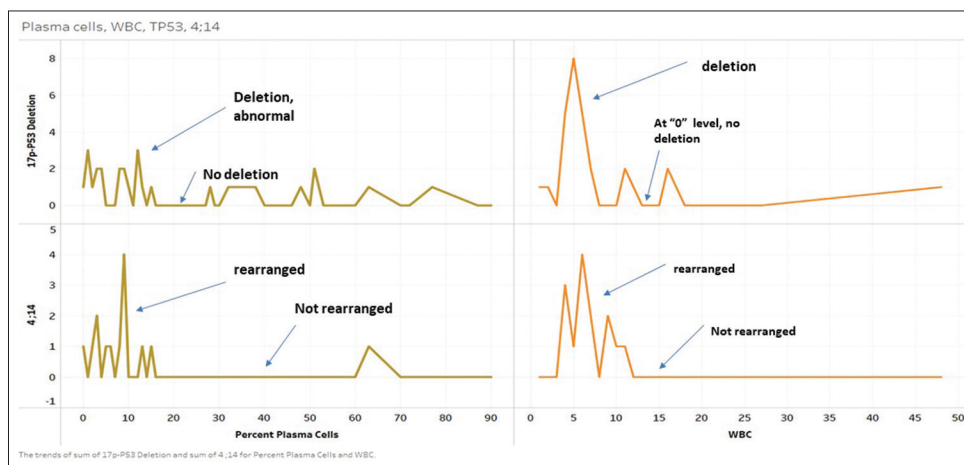

Figure 7: Tableau Cohort 1 data: TP53 and t(4;14) versus plasma call percent and white blood cell at zero-no deletion, peaks are number of cases with deletion
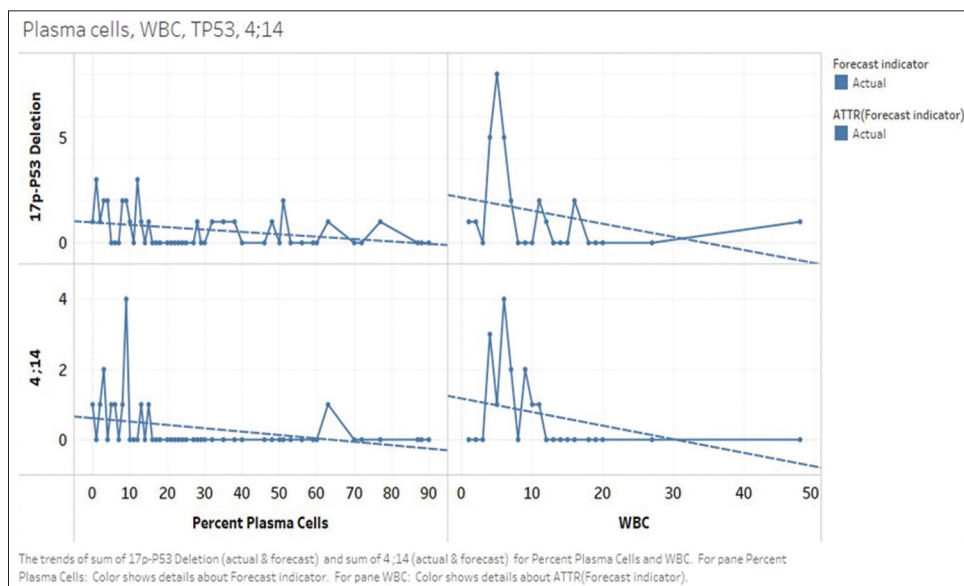

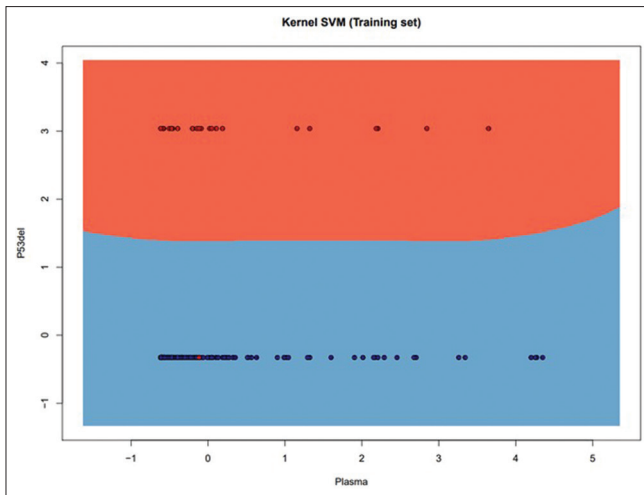**Figure 8:** Tableau Cohort 1 data – Forecast indicator

Figure 9: Support vector machine with MM4 data of cohort 1 between TP53 and plasma cell percentage, training 75% and test 25%
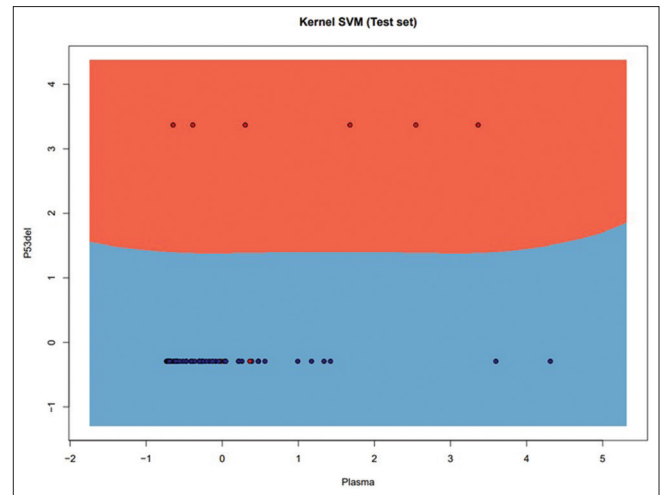

Figure 10: Support vector machine with MM4 data of cohort 1 between TP53 and plasma cell percentage, training 75%, test 25%
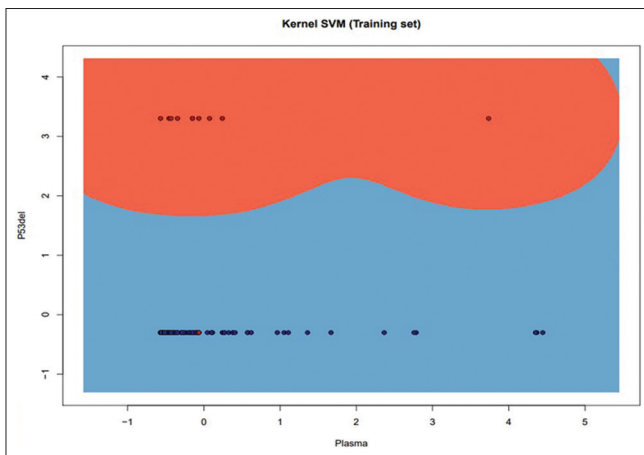

Figure 11: Support vector machine with MM4 data of cohort 1 between TP53 and plasma cell percentage, training 60%, test 40%
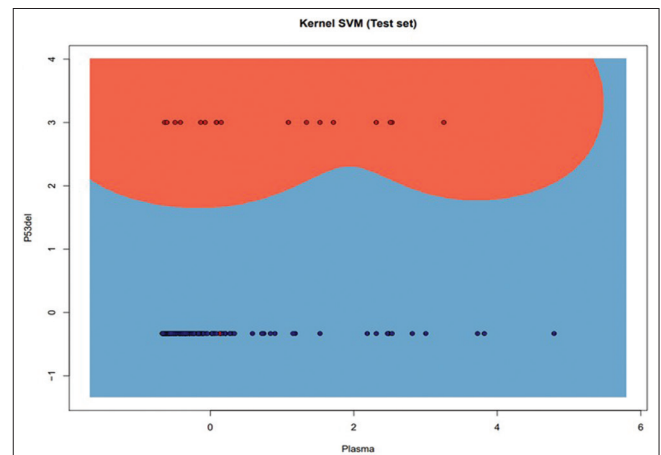

Figure 12: Support vector machine with MM4 data of cohort 1 between TP53 and plasma cell percentage, training 60% and test 40%

2. If so, could a classification model aid the critical variables in the disease management initiatives?

Response: Although preliminary, the models by ANN seem to identify that lower percentage of BM plasma cells and lower WBCs may be indicative of the two high-risk genetic abnormalities described above.

3. Would such an approach represent a promising tool for diagnosis and/or follow-up of PCM patients?

Response: If the model proves to be accurate, additional studies fitting the criteria of high-risk categories could be examined.

CONCLUSION

Morphology of BM and the diagnosis of genetic abnormalities by FISH, microarray, and/or DNA sequencing are critical for prognosis, risk stratification, selection of an appropriate therapeutic strategy, and persistent MRD by flow cytometry (NCCN) for PCM. Due to an increased interest in enrichment and downstream testing of PCM across the

clinical laboratories, research was initiated with a machine learning approach to identify if a relationship exists between different variables identified in the high-risk categories of PCM. Research articles published on machine learning in PCM are based mostly on clinical trials for predicting treatment benefit, multilevel drug response, ICD-9-CM diagnosis codes, administrative data using SEER registry, and distinguishing smoldering versus symptomatic multiple myeloma.[22-29] Application of AI techniques in PCM diagnosis is still at an early and unexplored stage.

Although the genetics of PCM is highly complex and heterogeneous with various clinical outcomes, an opportunity to categorize the disease into genetic subtypes for an effective treatment strategy is possible.[7-32] In this paper, we describe a novel classification model based only on specific morphological and genetic variables tested at our institution, using a machine learning algorithm, the ANNs. We identified a relationship of WBC and BM plasma cell percentage with two of the high-risk genetic categories at the

time of diagnosis in PCM. Visualization with Tableau analytics of the first data set for the high-risk categories precisely unveiled valuable insights into how and why the ANN worked with this data set. Our data analysis indicates that cases with high-risk abnormalities seem to be associated with low WBC and low BM plasma cell percentage. With further training and testing of additional data sets that include morphological and additional genetic rearrangements for high-risk category, ANNs have the potential to develop an accurate classification of high-risk categories of PCM at the time of diagnosis.

Exploration through ANN is not currently being utilized to assess the risk status of PCM patients at diagnosis. By uncovering these important diagnostic indicators, we believe that these insights can aid in early risk stratification, thus positively impacting the efficacy of future disease management. By discovering the correlation between WBC, BM plasma cell percentage, and genetic risk factors, actionable programs could be developed and targeted to high-risk groups. The tremendous potential of AI is that it can provide tools to explore massive data sets to identify the causal relationships of clinical findings, morphologic characteristics, and genetic alterations, thus being able to demonstrate useful and practical contributions to our knowledge in this domain.

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. LaCaria T., Schodel K. Case 411 -- A Solitary Bone Lesion [Internet]. University of Pittsburgh Department of Pathology; c2012 [updated 2004 Dec; cited 2021 Aug 04]. Available from: path.upmc.edu/cases/case411.html.
2. Mikhael J, Ismaila N, Cheung MC, Costello C, Dhodapkar MV, Kumar S, *et al.* Treatment of multiple myeloma: ASCO and CCO joint clinical practice guideline. J Clin Oncol 2019;37:1228-63.
3. NCCN.org [Internet]. Washington, D.C.: National Comprehensive Cancer Network, c2021 [cited 2020 Apr]. Available from: https://NCCN.org/store/login/login.aspx?ReturnURL=https://www.nccn.org/professionals/physician_gls/pdf/myeloma.pdf.
4. Cancer.org [Internet]. Washington, D.C.: American Cancer Society, c2021 [cited 2020 Apr]. Available from: https://cancer.org/cancer/multiple-myeloma.html.
5. National Institute of Health (NIH). //www.nih.gov.
6. SEER Cancer Statistics: 1975-2016. https://seer.cancer.gov/faststats.
7. Munshi NC, Anderson KC, Bergsagel PL, Shaughnessy J, Palumbo A, Durie B, *et al.* Consensus recommendations for risk stratification in multiple myeloma: Report of the International Myeloma Workshop Consensus Panel 2. Blood 2011;117:4696-700.
8. "Munshi NC, Avet-Loisseau H, Rawstron AC, Association of minimal residual disease with superior survival outcomes in patients with multiple myeloma: A meta-analysis. JAMA Oncol. 2017 Jan 1; 3(1): 28-35. doi: 10.1001/jamaoncol.2016.3160.
9. Kumar S, Paiva B, Anderson KC, Durie B, Landgren O, Moreau P, *et al.* International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol 2016;17:e328-46.
10. Perrot A, Corre J, Avet-Loiseau H. Risk stratification and targets in multiple myeloma: From genomics to the bedside. Am Soc Clin Oncol Educ Book 2018;38:675-80.
11. Anderson KC, Alsina M, Atanackovic D, Biermann JS, Chandler JC, Costello C, *et al.* Multiple myeloma, version 2.2016: Clinical practice guidelines in oncology. J Natl Compr Canc Netw 2015;13:1398-435.
12. Morgan G, Walker B, Davies F. The genetic architecture of multiple myeloma. Nat Rev Cancer 2012;12:335-48.
13. Talley PJ, Chantry AD, Buckle CH. Genetics in myeloma: Genetic technologies and their application to screening approaches in myeloma. Br Med Bull 2015;113:15-30.
14. Bolli N, Maura F, Minvielle S, Gloznik D, Szalat R, Fullam A, *et al.* Genomic patterns of progression in smoldering multiple myeloma. Nat Commun 2018;9:3363.
15. Corre J, Munshi N, Avet-Loiseau H. Genetics of multiple myeloma: Another heterogeneity level? Blood 2015;125:1870-6.
16. Walker BA, Wardell CP, Melchor L, Brioli A, Johnson DC, Kaiser MF, *et al.* Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. Leukemia 2014;28:384-90.
17. Pugh TJ, Fink JM, Lu X, Mathew S, Murata-Collins J, Willem P, *et al.* Assessing genome-wide copy number aberrations and copy-neutral loss-of-heterozygosity as best practice: An evidence-based review from the Cancer Genomics Consortium working group for plasma cell disorders. Cancer Genet 2018;228-229:184-96.
18. Jung HA, Jang MA, Kim K, Kim SH. Clinical utility of a diagnostic approach to detect genetic abnormalities in multiple myeloma: A single institution experience. Ann Lab Med 2018;38:196-203.
19. Hebraud B, Magrangeas F, Cleynen A, Lauwers-Cances V, Chretien ML, Hulin C, *et al.* Role of additional chromosomal changes in the prognostic value of t(4;14) and del(17p) in multiple myeloma: The IFM experience. Blood 2015;125:2095-100.
20. Keen-Kim D, Kaplan L, Siva A, Wolfe K, Nooraie F, Chan AC, *et al.* Intelligent FISH for Myeloma: Enhanced Cytogenetic Aberration Detection with Immunomagnetic Bead Enrichment of Plasma Cells. Myeloma Biology and Pathophysiology, Excluding Therapy: Poster I; 2010. doi: 10.1182/blood.V116.21.1918.1918.
21. Joseph NS, Gentili S, Kaufman JL, Lonial S, Nooka AK. High-risk multiple myeloma: Definition and management. Clin Lymphoma Myeloma Leuk 2017;17S:S80-7.
22. Sanyal P, Mukherjee T, Barui S, Das A, Gangopadhyay P. Artificial intelligence in cytopathology: A neural network to identify papillary carcinoma on thyroid fine-needle aspiration cytology smears. J Pathol Inform 2018;9:43.
23. Walker BA, Boyle EM, Wardell CP, Murison A, Begum DB, Dahir NM, *et al.* Mutational spectrum, copy number changes, and outcome: Results of a sequencing study of patients with newly diagnosed myeloma. J Clin Oncol 2015;33:3911-20.
24. Deulofeu M, Kolářová L, Salvadó V, Peña-Méndez EM, Almáši M, Štork M, *et al.* Rapid discrimination of multiple myeloma patients by artificial neural networks coupled with mass spectrometry of peripheral blood plasma. Sci Rep 2019;9:1-7.
25. Mesko B. The role of artificial intelligence in medicine. Expert Review of Precision Medicine and Drug Development, 2017; 2:5, 239-241 [cited 2020 Apr]. Available from: https://www.tandfonline.com/doi/full/10.1080/23808993.2017.1.380516.
26. Basheer IA, Hajmeer M. Artificial neural networks: Fundamentals, computing, design, and application. J Microbiol Methods 2000;43:3-31.
27. Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. J Appl Biomed 2013;11:47-58.
28. Ardizzone E, Bonadonna F, Gaglio S, Marcenò R, Nicolini C, Ruggiero C, *et al.* Artificial intelligence techniques for cancer treatment planning. Med Inform (Lond) 1988;13:199-210.
29. Deulofeu M, Kolářová L, Salvadó V, María Peña-Méndez E, Almáši M, Štork M, *et al.* Rapid discrimination of multiple myeloma patients by artificial neural networks coupled with mass spectrometry of peripheral blood plasma. Sci Rep 2019;9:1-7.
30. Tensorflow.org [Internet]. [cited 2020 Apr]. Available from: https://tensorflow.org.
31. Keras.io [Internet]. [cited 2020 Apr]. Available from: https://keras.io.
32. Rajkumar SV. Multiple myeloma: 2020 update on diagnosis, risk-stratification and management. Am J Hematol 2020;95:548-67.