phuse·connect
connect·share·advance
The Clinical Data Science Conference
EU 2024
10-13 November
Strasbourg Convention Centre


Paper AS09

Principal Component Analysis (PCA) - An Efficient Way of Variable Reduction Method and its Application in Oncology Trials

Pradeep Acharya, Ephicacy Lifescience Analytics, Bengaluru, India

ABSTRACT

In clinical studies, it is not feasible to use all variables in final analysis model even though they have significant influence on the outcome variable. The independent variables are usually correlated to each other which can introduce multicollinearity when the greater number of variables are included in the model. Application of Principal Component Analysis (PCA) on these variables can help solve such problems. PCA is a statistical method used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set which can then be used for final analysis.

This paper illustrates how to apply PCA using SAS to reduce the dimension of oncology data without information loss. A clinical dataset of higher dimension has been used for the demonstration and the efficiency of PCA in reducing the variable dimension has been discussed.

INTRODUCTION TO PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is one of the most popular statistical techniques used in various scientific fields to reduce the dimension of the dataset. The main idea of PCA is to decrease the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This will be achieved by transforming existing set of variables to a new set of variables, called the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all the original variables.

When there are substantial number of explanatory variables in any project which influences the primary outcome variable, then all those explanatory variables need to be considered for the primary statistical analysis. Some of these explanatory variables may highly influence the outcome variable whereas some of them may not. Use of enormous number of variables in any statistical model complicates the analysis, creates higher level of interactions, and may lead to incorrect results and interpretations. In such cases, the Principal Component Analysis method gives a calculative way of eliminating some of the less influencing explanatory variables without losing the evidence, thereby maintaining the transparency of all information. PCA breaks down variables into a subset of linearly independent principal components which may then be used in subsequent analyses to get more appropriate results.

STEPS TO CALCULATE PRINCIPAL COMPONENTS

There are five important steps in constructing the principal components. A brief description of each of them has been given in this section.

STANDARDIZATION

To perform PCA, the data should follow a normal distribution with a mean of zero and a standard deviation of one. Usually, the variables may or may not be normal in research, it is essential to standardize it. In this step, the mean values of the variables are calculated and subtracted from the original dataset so that each variable contributes equally to the analysis. This value is then divided by the standard deviation for each variable so that all variables use the same scale.

COVARIANCE MATRIX COMPUTATION

The covariance matrix summarizes the covariances associated with all pair combinations of the initial variables in the dataset. Computing the covariance matrix helps identify the relationships between the variables. The positive sign of the values in the matrix indicates the variables are positively correlated and increase or decrease at the same time. The negative sign of the values indicates that the variables are negatively correlated, meaning that one decreases while the other increases, and a value of zero indicates the variables are not related to each other.

COMPUTATION OF EIGEN VECTORS AND EIGEN VALUES OF THE COVARIANCE MATRIX

Eigenvectors and eigenvalues are the linear algebra concepts that we need to compute from the covariance matrix to determine the principal components of the data. The eigenvectors (also known as the principal components) represent the

directions of maximum variance in the data. The eigenvalues represent the amount of variance in each component. Ranking the eigenvectors by eigenvalue identifies the order of principal components.

There are two Principal Components identified namely (i) The First Principal Component (PC1) and (ii) the Second Principal Component (PC2).

The first principal component (PC1) is the direction in space along which the data points have the highest or most variance. The larger the variability captured in the first component, the larger the information retained from the original dataset. No other principal component can have a higher variability.

Second principal component (PC2) is calculated in the same way as PC1. PC2 accounts for the next highest variance in the dataset and must be uncorrelated with PC1. That is, PC2 must be orthogonal, that is perpendicular to PC1. This relationship can also be expressed as the correlation between PC1 and PC2 equals zero.

CREATE FEATURE VECTOR

Computing the eigenvectors and ordering them by their eigenvalues in descending order, allows to find the principal components in order of their significance. Here the decision is made to keep the principal components with higher significance and to omit the PCs with lesser significance. The Feature Vector (FC) is simply a matrix that has as columns the eigenvectors of the components that one decided to keep. This makes it the first step towards dimensionality reduction, because if we choose to keep only p eigenvectors (or components) out of n, the final data set will have only p dimensions.

TRANSFORMING THE DATA INTO NEW COORDINATE SYSTEM

Here, one must use the feature vector formed using the eigenvectors of the covariance matrix, to reorient the data from the original axes to the ones represented by the principal components. This creates new data, capturing most of the information but with fewer dimensions than the original dataset.

GENERAL FORM OF FORMULAT TO CALCULATE PRINCIPAL COMPONENTS

This is the general formula to compute scores on the first component extracted (created) in a principal component analysis. The second and the subsequent principal components will also be calculated in the same way with the respective coefficients.

$C_1 = b_{11}(X_1) + b_{12}(X_2) + ... + b_{1p}(X_p)$

Where,

C_1 = the subject's score on principal component 1 (the first component extracted)

b_{1p} = the regression coefficient (or weight) for observed variable p, as used in creating principal component 1

X_p = the subject's score on observed variable p.

APPLICATIONS OF PCA

Principal Component Analysis (PCA), due to its wide range of usability, applied in different fields listed here, but which are not limited to,

- Mainly used in the analysis of Microarray data
- To visualize multidimensional data, PCA is used in machine learning
- Used in healthcare data, or in clinical research, to explore the factors that are assumed to be very important in increasing the risk of any chronic disease.
- PCA helps to resize an image.
- Used in the analysis of stock data and forecasting data.
- Helps to analyze patterns when dealing with high-dimensional data sets

METHODOLOGY, SOURCE DATA AND ITS BACKGROUND

This paper explains how PCA can be used in clinical research data. Firstly, an oncology study has been considered where the patients are treated for Lung Cancer. A hypothetical dataset of 216 patients has been generated which is identical to the actual clinical trial scenario where the primary outcome variable is Overall Survival (OS) of lung cancer patients. Two treatment groups have been created considering one active treatment and the other is placebo. The primary risk factors for lung cancers such as Number of Cigarettes Smoking per Day (NUMCIGD), Years Since Smoking (SMKDUR), Indirect Smoking in Years (INSMKDUR), Years of Exposure to Radon (EXPRADN), Years of Exposure to Asbestos (EXPASB), and the Body Mass Index (BMI) of the subjects have been created in the analysis dataset and have been considered for the final analysis to assess the efficacy of the treatment.

During univariate analysis, it is found that all these variables are significantly related to Overall Survival (OS) time of patients. Hence all these variables have been considered as explanatory variables for the final analysis. This simulated dataset as in *Image 1* has been used as the source data through out in this paper for the demonstration.

To illustrate the advantages of use of Principal Component Analysis, firstly the Cox Proportional Hazards model has been applied to analyze the Overall Survival time by including all the original explanatory variables mentioned above and the results are displayed. Then the PCA method has been used to reduce the dimension of the dataset by reducing the less contributing variables and created the linear combination of new variables called Principal Components which are

uncorrelated with each other. Then again Cox Proportional hazards model has been applied by using the principal components instead of original variables. The results obtained with the principal components have been compared with the results obtained with the original variables and the conclusion is given. All the analyses have been performed using SAS version 9.4.

usubjid	trt01pn	age	param	paramcd	aval	cnsr	NUMCIGD	SMKDUR	INSMKDUR	EXPRADN	EXPASB
01-701-1015	1	63	Overall Survival	OS	2	0	16	15.2	11.5	7.7	9.5
01-701-1023	1	64	Overall Survival	OS	3	0	14	2.3	4	6.9	11.6
01-701-1028	0	71	Overall Survival	OS	3	0	16	15	11.4	6.5	11.9
01-701-1033	0	74	Overall Survival	OS	28	1	12	1.2	3.4	7.2	10.9
01-701-1034	0	77	Overall Survival	OS	58	0	9	15.3	12.5	7.7	9.9
01-701-1047	1	85	Overall Survival	OS	46	1	8	2.2	1.3	6.8	11.6
01-701-1097	0	68	Overall Survival	OS	3	0	18	15.8	7.6	7.6	10.4
01-701-1111	0	81	Overall Survival	OS	11	1	22	0.8	2.6	8.2	9.1
01-701-1115	0	84	Overall Survival	OS	3	0	12	4.6	3.9	8.6	9.1
01-701-1118	1	52	Overall Survival	OS	182	1	14	15.2	11.5	9.1	8.3
01-701-1130	1	84	Overall Survival	OS	97	0	12	15.3	11.5	7.5	10.5
01-701-1133	0	81	Overall Survival	OS	61	0	10	15.3	12.6	7.4	10.7
01-701-1146	0	75	Overall Survival	OS	13	0	16	3.2	0.3	7.8	9.5
01-701-1148	0	57	Overall Survival	OS	3	0	15	15.2	12.5	7.4	10.8
01-701-1153	1	79	Overall Survival	OS	191	1	6	14.6	6.3	8.2	9
01-701-1180	0	56	Overall Survival	OS	36	0	16	2.9	1.4	7.5	10.5

Image 1: Dummy efficacy dataset

STATISTICAL ANALYSIS AND RESULTS

UNIVARIATE ANALYSIS TO CHECK THE INFLUENCE OF EXPLANATORY VARIABLES ON OUTCOME

Univariate analysis has been performed to see the relation between the explanatory variables and the Overall Survival Time. PROC CORR procedure (*Image 2*) is used to derive the Pearson's correlation coefficient and corresponding p values. According to the first row as highlighted in the Image 3, It is found that all the explanatory variables are significantly correlated ($p < 0.05$) with the Overall survival time of the patients. The *Image 3* shows the correlation coefficients and its respective p values for all combination of variables.

```
proc corr data=sltte;
  var aval numcigd smkdur insmkdur expasb bmibl;
run;
```

Image 2: SAS syntax to derive Pearson's correlation coefficients

Pearson Correlation Coefficients, N = 216 Prob > |r| under H0: Rho=0						
	aval	NUMCIGD	SMKDUR	INSMKDUR	EXPASB	bmibl
aval	1.00000	-0.19022 0.0050	0.53308 <.0001	0.44824 <.0001	-0.14803 0.0296	-0.14930 0.0282
NUMCIGD	-0.19022 0.0050	1.00000	-0.06486 0.3428	-0.02538 0.7107	-0.07502 0.2723	-0.07444 0.2761
SMKDUR	0.53308 <.0001	-0.06486 0.3428	1.00000	0.86267 <.0001	0.00888 0.8968	0.00856 0.9005
INSMKDUR	0.44824 <.0001	-0.02538 0.7107	0.86267 <.0001	1.00000	0.01343 0.8445	0.01338 0.8450
EXPASB	-0.14803 0.0296	-0.07502 0.2723	0.00888 0.8968	0.01343 0.8445	1.00000	0.99964 <.0001
bmibl	-0.14930 0.0282	-0.07444 0.2761	0.00856 0.9005	0.01338 0.8450	0.99964 <.0001	1.00000

Image 3: Pearson's correlation coefficients with p values

COX PROPORTIONAL HAZARDS MODEL WITH ALL ORIGINAL VARIABLES

To analyze the effect of treatments on the Overall Survival of the patients, Cox proportional hazards model has been applied with all the original explanatory variables. SAS syntax PROC PHREG as in *Image 4* has been written to get the proportional hazards model results. The variables used as independent variables along with treatment are Number of Cigarettes Smoking

per Day (NUMCIGD), Years Since Smoking (SMKDUR), Indirect Smoking in Years (INSMKDUR), Years of Exposure to Radon (EXPRADN), Years of Exposure to Asbestos (EXPASB), and the Body Mass Index (BMI) of the subjects.

```
proc phreg data=sltte;
  class trt01pn(ref="0");
  model aval*cnsr(1)=trt01pn numcigd smkdur insmkdur expasb bmibl/
                              ties=breslow risklimits alpha=0.05;
run;
```

Image 4: SAS syntax for Cox Proportional Hazards model with original explanatory variables

The overall model is non-significant with the Likelihood ratio test statistic 8.91 and the corresponding p value is 0.1789 as in the *Image 5*. Here five variables have been used in the independent list along with the treatment variable. Use of higher number of variables in any statistical model can result in multicollinearity, reduction in power of the test and hence non-accurate interpretation of the results. Here the non-significance of the overall model might have occurred due to multi-collinearity between the variables.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	8.9068	6	0.1789

Image 5: Model fit test for Cox Proportional Hazards model with original explanatory variables

Even though all the explanatory variables are significant contributors based on the univariate test results as in *Image 3*, the final overall model shows none of the variable is influencing the outcome variable. As we discussed earlier, this might have been caused due to the inter-relationship between the explanatory variables. The Hazard ratio corresponding Treatment variable is 0.699 (second highlighted box in *Image 6*) with 95% confidence interval is (0.409, 1.196), which represents the risk of getting an event in treatment group is approximately 30% times lesser than that of Placebo group. Here the confidence interval contains the value 1, which represents, there could be a situation where the number of events in active treatment group and the number of events in placebo are same. Hence, the actual or accurate inference of the treatment's efficacy is failed here.

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
trt01pn	1	1	1.7100	0.1910	0.699	0.409	1.196
NUMCIGD		1	1.2890	0.2562	1.035	0.976	1.097
SMKDUR		1	1.1561	0.2823	1.043	0.966	1.127
INSMKDUR		1	0.0784	0.7794	0.988	0.908	1.075
EXPASB		1	0.0497	0.8235	0.479	0.001	10.515
bmibl		1	0.0682	0.7940	1.390	0.117	16.481

Image 6: Estimates and HR of CPH model with original explanatory variables

PRINCIPAL COMPONENT ANALYSIS FOR REDUCTION OF VARIABLES

The procedure PROC PRINCOMP as in *Image 7* has been used to perform PCA to reduce the number of variables without losing the information.

```
proc princomp data=sltte out=sltte1 outstat=stat;
  var numcigd smkdur insmkdur expasb bmibl;
run;
```

Image 7: SAS syntax for Principal Component Analysis

The PCA method creates linear combination of variables called Principal Components (PC), and the PCs are selected based on the Eigen values. The procedure creates the same number PCs as many numbers of variables used in the procedure. The components with eigen values greater than 1 (or remarkably close to 1) need to be considered for the final analysis since the largest variance is explained by those components only. In *Image 8*, there are only two components which are having eigen value greater than 1 and one component is close to 1. The column proportion in this table explains that the percentage of variance explained by each of the components. The first component explains 40% of the variance, the second component explains 37% of the variance and the third component explains 19.7% of the variance. Altogether these three

components explain 97.3% of the variance in the data, and the rest two components explain very negligible variance, hence those can be neglected from the analysis.

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.01616243	0.15469869	0.4032	0.4032
2	1.86146374	0.87584263	0.3723	0.7755
3	0.98562111	0.84923061	0.1971	0.9726
4	0.13639050	0.13602828	0.0273	0.9999
5	0.00036222		0.0001	1.0000

Image 8: Eigen Values of the Correlation Matrix

The table in the *Image 9* has been used to create the first, second and the subsequent Principal Components. The first Principal Component (PC1) is calculated in this way.

PC_1 = -0.11*NUMCIGD + 0.13*SMKDUR + 0.13*INSMKDUR + 0.69*EXPASB + 0.69*BMIBL

Similarly, the second and the subsequent PCs are derived.

Eigenvectors					
	Prin1	Prin2	Prin3	Prin4	Prin5
NUMCIGD	-.113237	-.048797	0.991822	0.032928	0.000381
SMKDUR	0.132191	0.693725	0.025734	0.707537	-.000709
INSMKDUR	0.133253	0.691861	0.072688	-.705895	0.000591
EXPASB	0.689936	-.137207	0.071625	0.003729	0.707120
bmibl	0.689880	-.137356	0.072197	0.002448	-.707093

Image 9: Eigen Vector

The *out*= option in the syntax in *Image 7*, creates a dataset with the Principal Components of linear combination of all the independent variables. These PCs are independent of each other and have zero mean and constant variance. The Principal Components created by above procedure are Prin1, Prin2, Prin3, Prin4 and Prin5 as in the below *Image 10*. Based on the results from *Image 8*, the first three PCs only explain 97.3% of the variance in the data the rest two can be omitted from the analysis. Similarly, when we have a higher number of variables, this PCA effectively helps to reduce the data dimension.

usubjid	trt01pn	age	param	paramcd	aval	cnsr	NUMCIGD	SMKDUR	INSMKDUR	EXPRADN	EXPASB	Prin1	Prin2	Prin3	Prin4	Prin5
01-701-1015	1	63	Overall Survival	OS	2	0	16	15.2	11.5	7.7	9.5	0.3747015425	1.1427247578	1.0232703775	0.1240136499	-0.025145999
01-701-1023	1	64	Overall Survival	OS	3	0	14	2.3	4	6.9	11.6	2.5183452079	-1.915542857	0.5417854887	-0.377918539	0.0246649079
01-701-1028	0	71	Overall Survival	OS	3	0	16	15	11.4	6.5	11.9	3.3624495108	0.5091123945	1.3330977132	0.1274422089	-0.027936522
01-701-1033	0	74	Overall Survival	OS	28	1	12	1.2	3.4	7.2	10.9	1.7268832246	-1.938412452	-0.139734144	-0.448732417	-0.032592121
01-701-1034	0	77	Overall Survival	OS	58	0	9	15.3	12.5	7.7	9.9	1.123938024	1.2973583417	-0.950509781	-0.071616607	-0.014097534
01-701-1047	1	85	Overall Survival	OS	46	1	8	2.2	1.3	6.8	11.6	2.6432060652	-2.217712516	-1.24646049	-0.064103305	0.023684168
01-701-1097	0	68	Overall Survival	OS	3	0	18	15.8	7.6	7.6	10.4	1.3014840944	0.425866029	1.664576658	0.7754413083	0.0131142704
01-701-1111	0	81	Overall Survival	OS	11	1	22	0.8	2.6	8.2	9.1	-0.924151015	-1.785468336	2.5220870943	-0.29682662	0.0132548893
01-701-1115	0	84	Overall Survival	OS	3	0	12	4.6	3.9	8.6	9.1	-0.470691054	-1.010160042	-0.355937381	-0.11869536	0.0118299201
01-701-1118	1	52	Overall Survival	OS	182	1	14	15.2	11.5	9.1	8.3	-1.068136591	1.471739661	0.2831833061	0.0979220597	-0.011776278
01-701-1130	1	84	Overall Survival	OS	97	0	12	15.3	11.5	7.5	10.5	1.7281882159	0.9695378636	-0.0148429	0.1028973858	0.0037066685
01-701-1133	0	81	Overall Survival	OS	61	0	10	15.3	12.6	7.4	10.7	2.0678826434	1.1030471206	-0.556099752	-0.071779836	0.0094762491
01-701-1146	0	75	Overall Survival	OS	13	0	16	3.2	0.3	7.8	9.5	-0.244780041	-1.832290992	0.801482863	0.2647548653	0.0237437608
01-701-1148	0	57	Overall Survival	OS	3	0	15	15.2	12.5	7.4	10.8	2.006524012	0.9836979659	0.9103378602	-0.020783997	0.0251594211
01-701-1153	1	79	Overall Survival	OS	191	1	6	14.6	6.3	8.2	9	-0.066757748	0.6137056075	-2.034021234	0.6913750236	-0.029275467
01-701-1180	0	56	Overall Survival	OS	36	0	16	2.9	1.4	7.5	10.5	1.0437079382	-1.96642623	0.9482092406	0.0776825638	0.0044549962
01-701-1181	0	79	Overall Survival	OS	8	1	15	0.4	4.8	7.4	10.6	1.2216577403	-1.796627141	0.7070686712	-0.717440491	0.0201741014
01-701-1192	0	80	Overall Survival	OS	17	0	12	15.3	11.9	7.3	10.6	1.8488465547	1.003418417	0.0024654995	0.0465209498	0.0188776476
01-701-1203	1	81	Overall Survival	OS	183	1	13	15.3	13.5	7.8	9.8	0.9077944266	1.4012648072	0.2181951369	-0.175633637	-0.028654215
01-701-1234	1	69	Overall Survival	OS	177	1	18	14.8	8.1	8.7	8.9	-0.570621399	0.7476269784	1.4733243563	0.5750097537	0.0059026659
01-701-1275	0	61	Overall Survival	OS	18	0	12	9.5	7.1	8.2	9.3	0.0171815671	-0.04085842	-0.257146222	0.0203266919	-0.03128906
01-701-1287	0	56	Overall Survival	OS	29	0	14	15.3	11	8.2	8.9	-0.324612527	1.2637407887	0.3549670453	0.1844953085	-0.018644686

Image 10: Dataset with Principal Components as a linear combination of original variables

COX PROPORTIONAL HAZARDS MODEL WITH PRINCIPAL COMPONENTS AS EXPLANATORY VARIABLES

After the creation of Principal Components, the cox proportional model is done using the highest variance explained PCs instead of the original variables. Same SAS syntax as in *Image 4* has been used replacing the original variables with the Principal Components as explanatory variables. *Image 11* is the updated syntax for CPH model with PCs.

```
proc phreg data=sltte1;
  class trt01pn(ref="0");
  model aval*cnsr(1)=trt01pn prin1 prin2 prin3/ties=breslow risklimits alpha=0.05;
run;
```

Image 11: Model fit test for Cox Proportional Hazards model with PCs as explanatory variables

In the above procedure only the first three PCs (Prin1, Prin2 and Prin3) have been used in the model statement since these are three components only explain the highest variance in the data as well as it nullifies the inter-relationship of the explanatory variables since the PC are independent and standardized. *Image 12* shows the overall model turned to slightly significant with Likelihood ratio test statistic 19.78 and the p value is 0.0006.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	19.7808	4	0.0006

Image 12: Model fit test for Cox Proportional Hazards model with PCs as explanatory variables

The results in *Image 13* shows, that use of PCs in have helped to get the accurate results for the treatment and the treatment effect has appeared to be significant with the Chi Square value 5.25 and the p value 0.0219. The Hazard ratio corresponding Treatment variable also has decreased when compared with the original variables, and currently it is 0.324 with 95% confidence interval is (0.281, 0.808), which represents the risk of getting an event in treatment group is approximately 68% times lesser than that of Placebo group.

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Chi-Square	Pr > ChiSq	Hazard Ratio	95% Hazard Ratio Confidence Limits	
trt01pn	1	1	5.2503	0.0219	0.324	0.281	0.808
Prin1		1	12.2126	0.0005	1.227	1.084	1.389
Prin2		1	2.7872	0.0950	1.116	0.959	1.298
Prin3		1	0.9653	0.3259	1.118	0.915	1.367

Image 13: Estimates and HR of CPH model with PCs as explanatory variables

ADVANTAGES AND DISADVANTAGES OF PRINCIPAL COMPONENT ANALYSIS

Advantages of Principal Component Analysis

- Easy to calculate and compute.
- Speeds up machine learning computing processes and algorithms.
- Prevents predictive algorithms from data overfitting issues.
- Increases performance of ML algorithms by eliminating unnecessary correlated variables.
- Principal Component Analysis results in high variance and increases visualization.
- Helps reduce noise that cannot be ignored automatically.

Disadvantages of Principal Component Analysis

- Sometimes, PCA is difficult to interpret. In rare cases, you may feel difficult to identify the most important features even after computing the principal components.
- You may face some difficulties in calculating the covariances and covariance matrices.
- Sometimes, the computed principal components can be more difficult to read rather than the original set of components.
- The interpretation is not on original variables, instead it is on calculated linear combination of the original variables

CONCLUSION

By looking at the difference in results obtained from traditional CPH model and CPH model with PCs as explanatory variables, it is always better to do the Principal Component Analysis (PCA) when there are vast number of variables which are exceedingly difficult to analyze and interpret. PCA is one of the powerful methods to reduce the data dimension without losing the information. Also, the limited number of PCs will only account for maximum variance in the data, it is easy to use, analyze and interpret the results. Also, the results obtained are more dependable since the highest variance is accounted for by these PCs.

REFERENCES

[1] A Ghosha, S Barmanb. Prediction of Prostate Cancer cells Based on Principal Component Analysis Technique. Procedia Technology 10 (2013) 37-44

[2] Alexe G, Dalgin G S, Ganesan S, DeLisi C and Bhanot G. Analysis of breast cancer progression using principal component analysis and clustering; J. Biosci. 32 (2007) 1027–1039

[3] Ying-Lin Hsu, Po-Yu Huang, Dung-Tsa Chen. Sparse principal component analysis in cancer research. Transl. Cancer Res. 2014;3(3):182-190

[4] T Bankole-Oye, I El-Thalji and J Zec. Combined principal component analysis and proportional hazard model for optimizing condition-based maintenance. IOP Conf. Series: Materials Science and Engineering 1201 (2021)

[5] I.T. Jolliffe. Principal Component Analysis, Second Edition. 2002, Springer Verlag, New York, Berlin, Heidelberg.

ACKNOWLEDGMENT

The author would like to extend his sincere thanks to Ephicacy Lifescience Analytics for the opportunity to write and present this paper. Any brand and product names are trademarks of their respective companies.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Pradeep Acharya
Ephicacy Lifescience Aalytics Pvt. Ltd. Bengaluru
Off Bannerghatta Main Road, Arekere, Bengaluru
080-41463195
Pradeep.acharya@ephicacy.com
www.ephicacy.com