MATHEMATICAL ANALYSIS AND PROBABILITY DENSITY FUNCTION OF FKHR PATHWAY FOR CELL SURVIVAL /DEATH

Shruti Jain

Department of Electronics and Communication, Jaypee University of Information Technology,
Waknaghat, Solan, Himachal Pradesh. 173234. India
jain.shruti15@gmail.com

ABSTRACT: This paper examines the mathematical analysis of the survival/ death decision using *tumor necrosis factor*, *epidermal growth factor* and *insulin* for *Forkhead transcriptional factors* (FKHR). The model was made using linear and non-linear methods. Regression analysis, KS test, AD test, chi square test, ANOVA, pdf includes the linear analysis while MLP and RBF methods using neural network constitute non- linear modeling. Regression analysis in which regression coefficient, adjusted regression coefficient, PRESS, regression coefficient cross validation, Durban Watson statistics were analyzed. We have also used different types of regression analysis to get different parameters. We have plotted the probability density function (pdf) of different functions for FKHR which is coming OK. We have also plotted the survival function and hazard function for different distribution functions.

KEYWORDS: Epidermal growth factor, Insulin, Mathematical modeling, Regression analysis, Distribution functions.

INTRODUCTION

Communication between the cell death / survival using different proteins was examined [1, 2, 3]. *Tumor necrosis factor-α* (TNF-α) was used as cell death protein [4, 5] while Epidermal growth factor (EGF) [6, 7, 8, 9] and insulin [10, 11, 12, 13] was used as cell survival protein. There are variations in the magnitude responses with cell type but pathways will remain the same [14, 15, 16].

Binding of Sh2 and ErbB leads to PI3K which is through EGF and insulin receptors [8, 9]. PI3K further leads to phosphorylates phosphatidylinositol (4, 5) biphosphate (PtdIns (4, 5)P_2) and RAS. PtdIns(3,4,5)P3 is a key protein to activate AkT. AkT leads to cell survival by activating Bad, NF-kB , CREB and leads to cell death by deactivating FKHR, GSK-3B. Table 1 shows the truth table for cell death/ survival for AkT/ FKHR pathway. AkT also activates mTOR. Figure 1 shows the communication pathways induced by combination of EGF and insulin leading to cell survival/ death.

1. EGF / PI3K / AkT → FKHR (Cell death)
2. Insulin / PI3K / AkT → FKHR (Cell death)

In this work our purpose is to determine whether the mathematical modeling using linear model and non- linear model can be used as important aspects of biological systems. Specifically we examine the TNF-α, EGF and insulin for cell survival/ death response of HT-29 human colon carcinoma cells based on measurements of 10 levels for FKHR. There are different types of linear modeling which we have used in this paper like : r^2, SER, PRESS, calculation of different parameters using different types of regression analysis, KS test, chi square test, AD test, ANOVA. Later, we are using different distribution functions (normal, weibull, logistic, lognormal base e, lognormal base 10, exponential) and plotted the probability function, survival function and hazard function. For non- linear approach we have used neural network in which different models based on MLP and RBF were used.

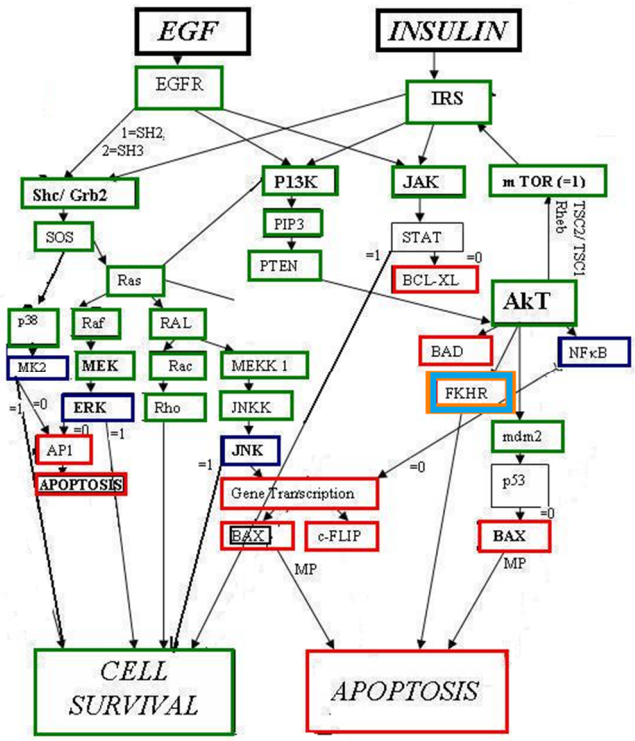

Fig 1 : Pathways of EGF/ Insulin for FKHR using AkT

SYSTEM IMPLEMENTATION USING LINEAR MODELING

Linear modeling was done by three methods: using regression analysis, different tests (KS, chi square, AD) on different functions and different plots (survival, hazard, probability) using different distribution functions.

Regression Analysis

We have designed a model using regression analysis for cell survival/ death. Different parameters are calculated with the different concentrations of the TNF, EGF and Insulin [2].

For r^2 , r^2_{pred}, r^2_{adj} :. The best equation for regression coefficient (r^2), adjusted regression coefficient (r^2_{adj}) and predicted regression coefficient(r^2_{pred}) was used. Equation 1 gives the r^2 equation.

$$r^2 = \frac{\sum\left(f_i - \bar{f}\right)^2}{\sum\left(y_i - \bar{y}\right)^2} \qquad (1)$$

where y_i and f_i are the observed and predicted values respectively while $\bar{y}$ and $\bar{f}$ are the means of corresponding values. In our case values are : S = 0.005784, r^2 = 92.6%, r^2 (adj) = 92.3%, r^2 (pred) = 92.03%.

We clubbed all the concentrations of TNF, EGF and Insulin and only normalized output (FKHR) was taken and we get the regression equation as:

Final Output = 0.557 + 0.000172a - 0.0000547b- 0.000083c + 0.000010d- 0.000033e - 0.000125f -0.000031g- 0.000173h+0.0000647i -0.000160j where a, b, c….. are the different concentrations of the three inputs.

We have also calculated : Mean sq error (MSE) , Root mean sq error (RMSE), Mean abs error (MAE), Relative sq error (RSE) , Root relative sq error (RRSE) and Relative abs error (RAE) for FKHR using different regression analysis like PLS, Linear, SVM, kNN, random forest, regression tree which was given in Table 1.

PRESS : The prediction sum of squares value should be small so as to get better model. In our case the PRESS value is coming out to be 0.010413.

The regression coefficient cross validation (q^2_{cv}), : For a perfect model q^2_{cv} should be close to one but it should be less than r^2 . For calculation of the q^2_{cv} we use equation 2:

Table 1 : Various analysis parameters using diff regression methods for FKHR

	MSE	RMSE	MAE	RSE	RRSE	RAE
PLS Regression	0.0000	0.0059	0.0047	0.0798	0.2825	0.2488
Linear Regression	0.0000	0.0059	0.0047	0.0798	0.2825	0.2488
SVM Regression	0.0005	0.0218	0.0209	1.0869	1.0426	1.1062
K nearest neighbors regression	0.0000	0.0069	0.0057	0.1102	0.3320	0.2999
Mean	0.0004	0.0210	0.0190	1.0100	1.0050	1.0046
Random Forest regression	0.0000	0.0064	0.0050	0.0951	0.3084	0.2668
Regression Tree	0.0000	0.0060	0.0048	0.0826	0.2874	0.2530

$$q^2_{cv} = 1 - \frac{PRESS}{TOTAL} = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \qquad (2)$$

In our case the q^2_{cv} value is coming out to be 0.20318.

Durbin Watson Statistics: d is approximately equal to $2(1 - r)$, where r is the sample autocorrelation of the residuals. If $d > 2$ indicated negatively correlation, $d = 2$ indicates no autocorrelation, $d < 2$ indicates positive serial correlation and if $d < 1$ causes alarm. The value of d always lies between 0 and 4. The Durbin-Watson statistic for FKHR is 1.98.

ANOVA : The analysis of the variance was shown in Table 2, which shows the sum of squares and mean squares of the regression and residual error.

Table 2 : Analysis of Variance for all combinations

Source	dF	Sum of sq	Mean of sq	F
Regression	10	0.121014	0.012101	361.71
Residual Error	289	0.009669	0.000033	
Total	299	0.130683		

Standard Error Coefficients (SER) : SER is used to measure precision. The smaller the value the more precise is the estimate. To get the *t*-value SER is divided by coefficient values. Table 3, shows the regression analysis in terms of SER, *t*-value, *p*-value and VIF.

Table 3: Regression analysis in terms of standard error coefficients, t-value, p value and

Predictor	Coef	SER	*t*-Value	*p*- Value	VIF
Constant	0.55687	0.03119	17.85	0.000	
0-0-0	0.00017162	0.00008008	2.14	0.033	69.0
5-0-0	0.00005418	0.00005264	1.03	0.304	5.5
100-0-0	-0.00008349	0.00002848	-2.93	0.004	131.5
0-100-0	0.00000983	0.00004522	0.22	0.828	6.8
5-1-0	0.00003322	0.00006344	0.52	0.601	25.1
100-100-0	-0.00012455	0.00004550	-2.74	0.007	29.2
0-0-500	-0.00003052	0.00005163	-0.59	0.555	7.1
0.2-0-1	-0.00017282	0.00005713	-3.03	0.003	143.3
5-0-5	0.00006400	0.00005365	1.19	0.234	98.5
100-0-500	-0.00016003	0.00004820	-3.32	0.001	118.9

Model using K-S test, Anderson darling test and chi-square test

There are different test performed on different distribution functions : Kolmogorov–Smirnov test, Anderson darling test and chi- square test.

The Kolmogorov–Smirnov (K-S/ KS) test is an equality test. One sample/one dimensional K-S test is used to compare a sample with a prob function while two sample/ 2-D test is used to compare two samples.

The Anderson Darling (AD) test/ Shapiro Wilk test is a statistical test and is based on the distance

$$A = n\int_{-\infty}^{\infty} \frac{\left(F_n(x) - F(x)\right)^2}{F(x)\left(1-F(x)\right)} dF(x) \qquad ..(3)$$

where $w(x) = \left[F(x)\left(1-F(x)\right)\right]^{-1}$ weight function.

A chi-squared test, also referred to as χ^2 test (or chi-square test), is any statistical hypothesis test which is used to determine whether there is a significant difference between the observed frequencies and the expected frequencies in one or more categories. Table 4 shows the KS, AD and chi square value of different functions.

Table 4: KS, AD and chi-square values of different distribution function

	K-S d	K-S	AD Stat	AD p-value	Chi-square	Chi-square p-value	Chi-square df
Gaussian Mixture(Mixing.Coef.1,Mean 1, Std.Dev 1, Mixing Coef.2,...)	0.026799	0.978501	0.1503	0.998562	6.067	0.108415	3
Weibull (scale,shape)	0.183078	0.000000	18.5049	0.000000	193.533	0.000000	7
Triangular(min,max,mode)	0.210974	0.000000	25.5090	0.000000	232.733	0.000000	6
Normal (location,scale)	0.231571	0.000000	22.6154	0.000000	266.733	0.000000	7
Log Normal (scale,shape)	0.236718	0.000000	23.2578	0.000000	279.467	0.000000	7
Rayleigh (scale)	0.568500	0.000000	117.2716	0.000000	1424.267	0.000000	8
Half Normal (scale)	0.640743	0.000000	144.1607	0.000000	1907.267	0.000000	8
General Pareto (scale,shape)	0.867025	0.000000	532.6983	0.000000	1555.267	0.000000	7

Distribution Functions

There are different distribution functions like Normal, Weibull, Lognormal base e, Lognormal base 10, Exponential, and Logistic.

A normal distribution is expressed as

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad ...(4)$$

where μ is the *mean* or *expectation* (median and mode) of the distance; σ is *standard deviation*. If $\mu = 0$ and $\sigma = 1$ the distance is called standard/unit normal distance. The probabilty density function (*pdf*), probability function, survival function and hazard function of normal distribution is shown in fig 2. We have calculated the Anderson darling adjustment values shown in Table 5.

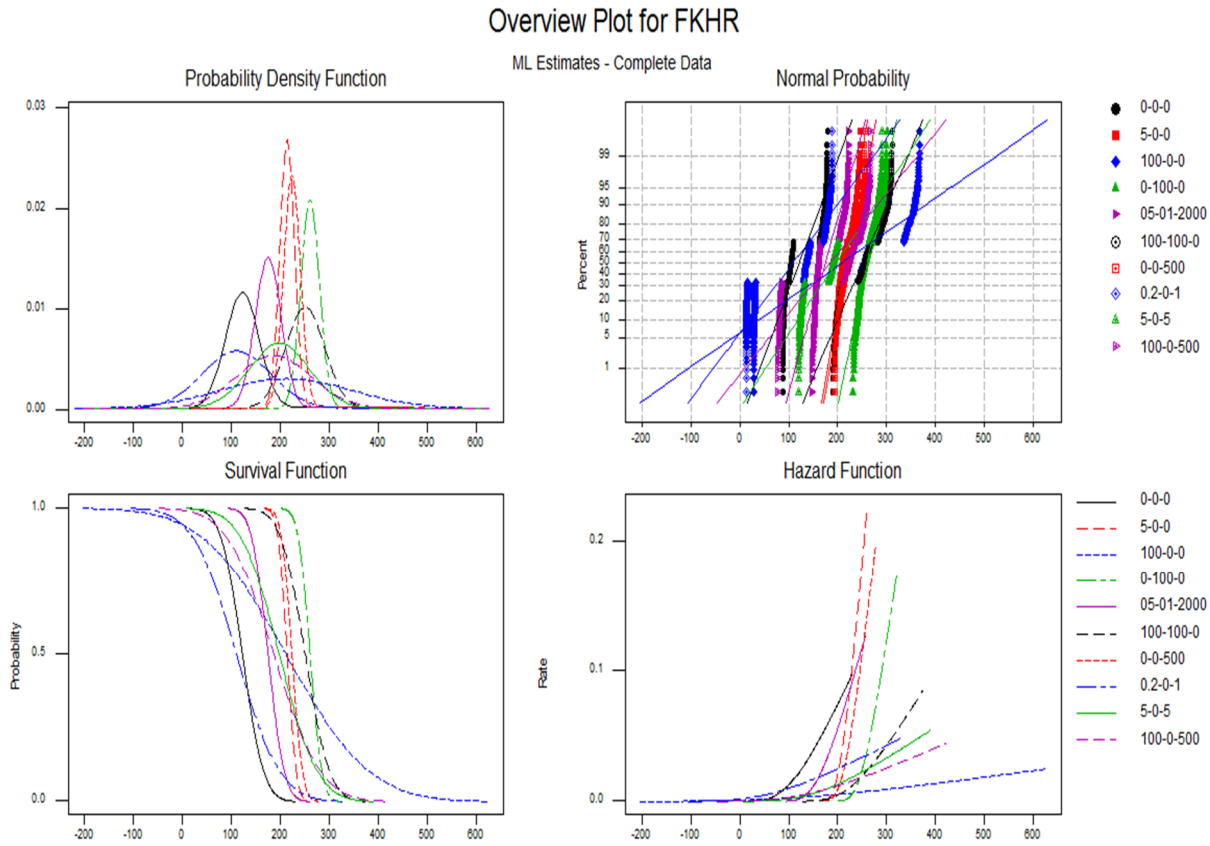

Fig 2 : pdf, probability, survival function and hazard function for Normal distribution

A Weibull distribution function : The probability density function of a Weibull function is expressed as

$$f(x;\lambda,k)=\begin{cases}\dfrac{k}{\lambda}\left(\dfrac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x\geq 0\\ 0 & x<0\end{cases}$$

..(5)

where $k > 0$ is the *shape parameter* and $\lambda > 0$ is the *scale parameter* of the distribution. If $k = 2$ and $\lambda = \sqrt{2}\,\lambda$, than weibull function equals to Rayleigh distribution. The probabilty density function (pdf), probability function, survival function and hazard function of weibull distribution is shown in fig 3. We have calculated the Anderson darling adjustment values shown in Table 5.

An exponential distribution function : In Equation 5 if $k = 1$ than weibull distribution equals to exponential distribution. The probability density function of an exponential distribution is expressed as

$$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0,\\ 0 & x<0.\end{cases}$$

…(6)

If $\lambda > 0$ is than the distribution, is called the *rate parameter*. The probabilty density function (pdf), probability function, survival function and hazard function of exponential distribution is shown in fig 4. We have calculated the Anderson darling adjustment values shown in Table 5.

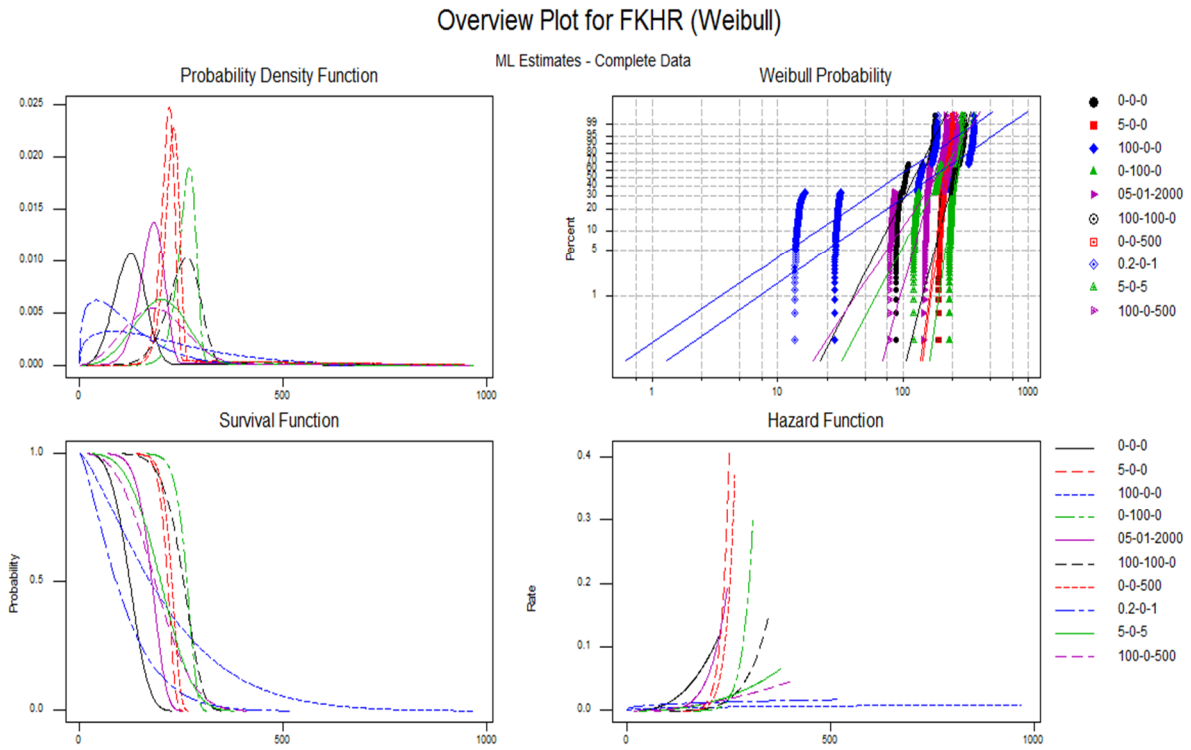

Fig 3: pdf, probability, survival function and hazard function for Weibull distribution

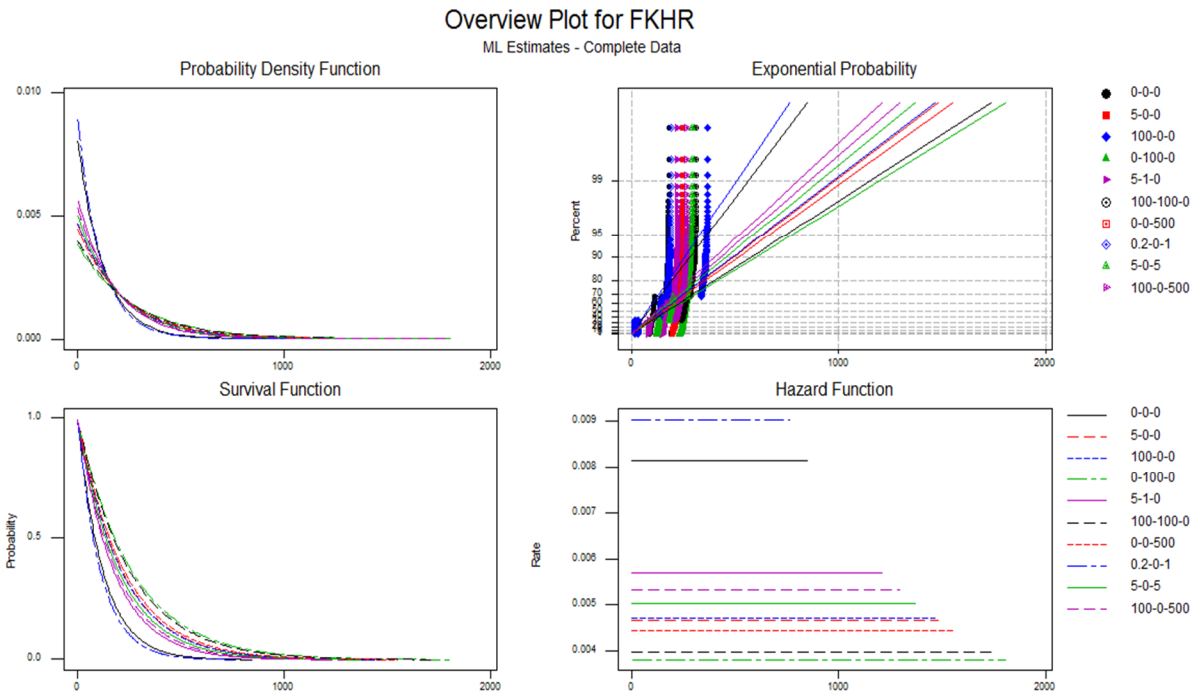

Fig 4 : pdf, probability, survival function and hazard function for exponential distribution

A lognormal distribution function : If the random variable X is log normally distributed then Y = log (X) is normal distribution function. Similarly if Y is a normal distribution than X = exp (Y) has a log normal distribution. The log normal function only takes real values. The probabilty density function (pdf), probability function, survival function and hazard function of log normal base e distribution and log normal base 10 distribution is shown in fig 5 and fig 6 respectively. We have calculated the Anderson darling adjustment values shown in Table 5.

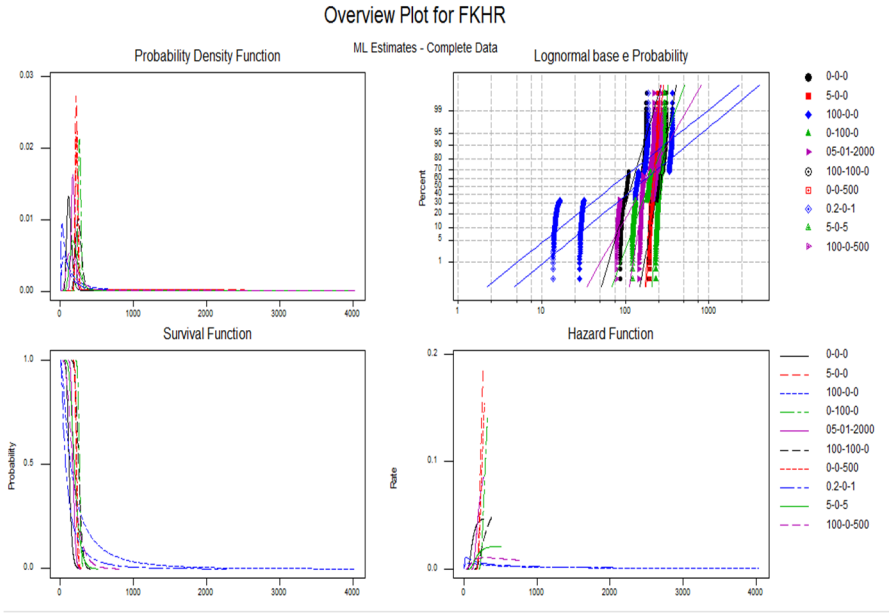

Fig 5: pdf, probability, survival function and hazard function for lognormal base e distribution

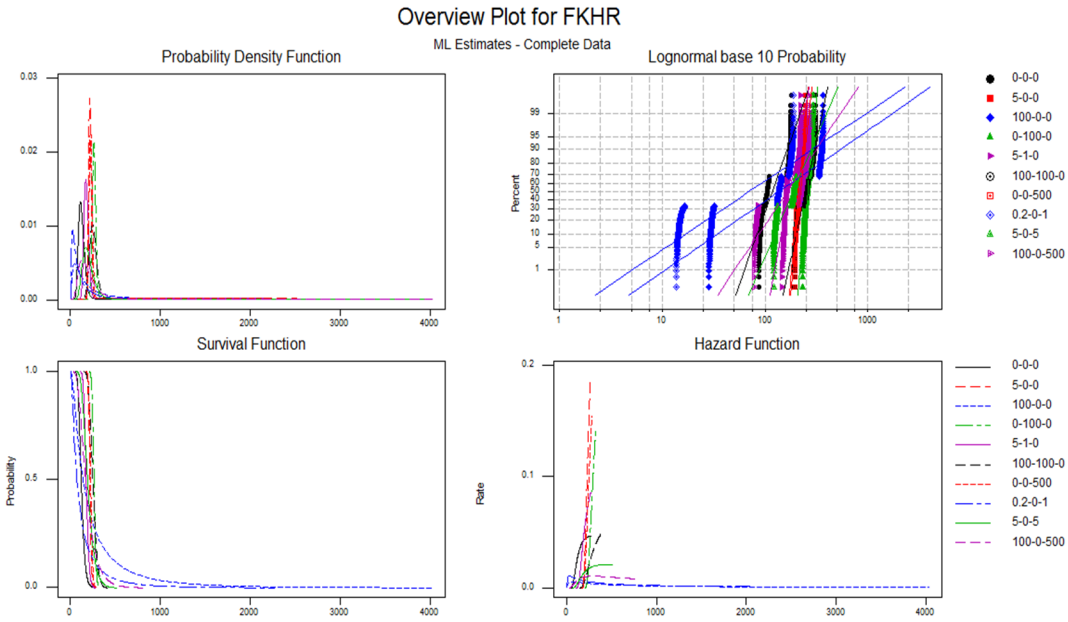

Fig 6 : pdf, probability, survival function and hazard function for lognormal base 10 distribution

A logistic distribution function/ sech- square distribution: The pdf of the logistic distribution is given by:

$$f(x;\mu,s)=\frac{e^{-\frac{x-\mu}{s}}}{s\left(1+e^{-\frac{x-\mu}{s}}\right)}=\frac{1}{4s}\sec h^2\left(\frac{x-\mu}{2s}\right) \qquad \ldots(7)$$

The probabilty density function (pdf), probability function, survival function and hazard function of logistic distribution is shown in fig 7. We have calculated the Anderson darling adjustment values shown in Table 5.

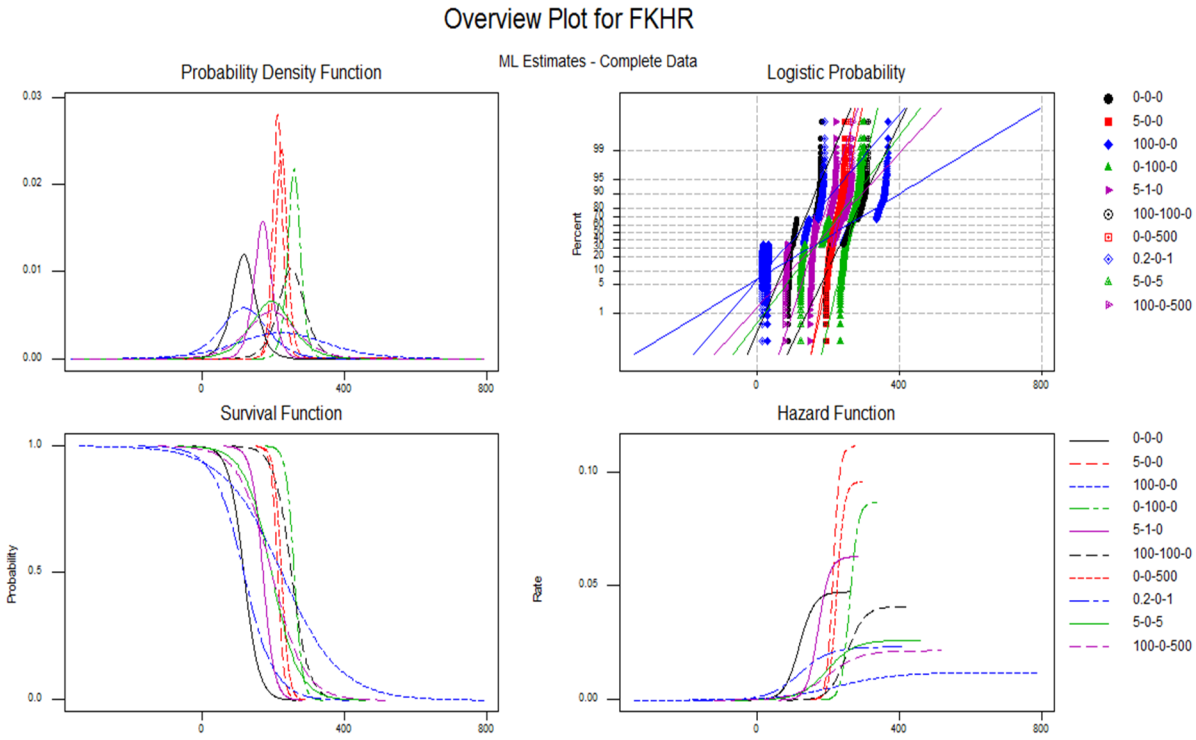

Fig 7: pdf, probability, survival function and hazard function for logistic distribution

Table 5 : The Anderson darling Adjustment values for different distributions

Predictor	Normal	Weibull	Exponential	Lognormal base 10	Lognormal base e	Logistic
0-0-0	33.02	30.76	78.4	28.19	28.19	30.50
5-0-0	6.14	8.60	119.1	5.38	5.38	6.06
100-0-0	27.41	42.47	34.9	44.06	44.06	24.67
0-100-0	7.04	9.71	118.2	6.10	6.10	6.84
5-1-0	32.78	31.39	103.5	30.83	30.83	30.65
100-100-0	10.03	8.90	97.8	11.48	11.48	9.48
0-0-500	2.81	3.90	117.0	2.69	2.69	3.20
0.2-0-1	30.31	45.62	37.4	45.89	45.89	27.28
5-0-5	16.78	16.29	67.3	16.21	16.21	15.52
100-0-500	35.37	41.49	55.4	43.02	43.02	31.99

All the above are parametric hazard functions. Fig 8 shows the non-parametric hazard function and Kaplan Meier Survival function.

SYSTEM IMPLEMENTATION USING NON-LINEAR MODELING

A neural network (ANN) model was developed for the AkT which predicts whether cell will survive or die. For training the ANN model [17, 18] experimental data from ten different concentrations of each marker proteins was taken as input, and their corresponding possible experimental output. We have implemented the Neural Network model using STATISTICA data miner software.

The training perfection, test perfection and validation perfection using MLP and RBF methods of 10 possible combinations for FKHR using automated neural search (ANS) and custom neural network (CNS) is shown in table 6 and table 7 respectively.

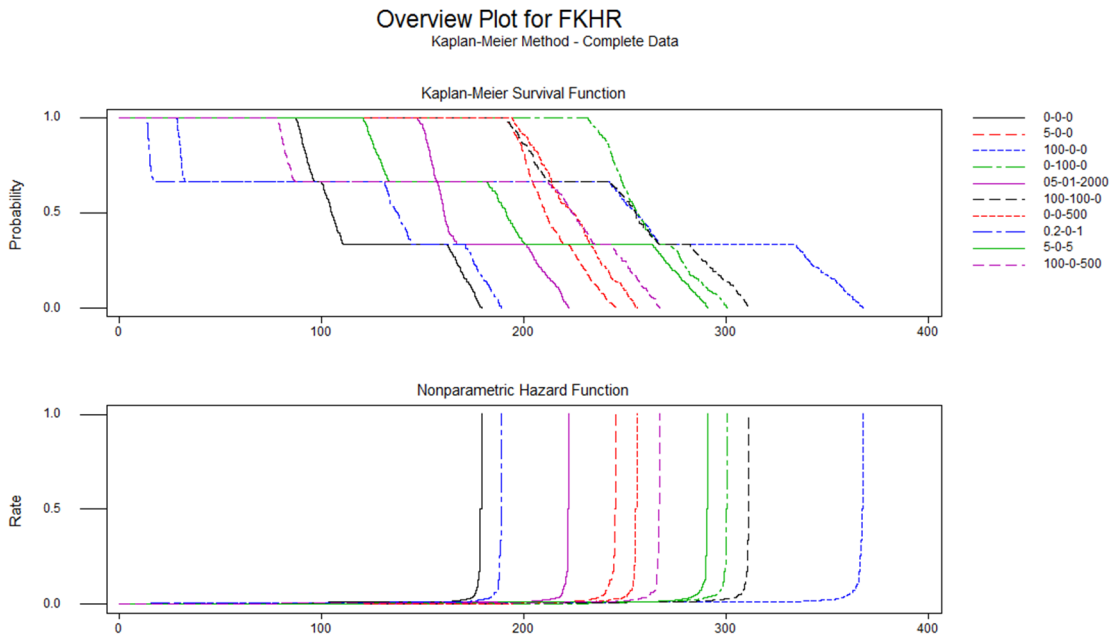

Fig 8 : The non-parametric Hazard function and Kaplan Meier Survival function

Table 6 : The output of FKHR for different concentration using ANS

S. No	Network Name	Training perfection	Test perfection	Validation perfection	Training algorithm	Hidden activation	Output activation
1	MLP 10-12-1	0.965069	0.959301	0.961561	BFGS 24	Tanh	Exponential
2	MLP 10-11-1	0.964972	0.960349	0.960810	BFGS 23	Tanh	Exponential
3	MLP 10-6-1	0.965628	0.962689	0.960205	BFGS 14	Exponential	Logistic
4	MLP 10-7-1	0.964834	0.960672	0.960744	BFGS 18	Exponential	Identity
5	MLP 10-10-1	0.962378	0.954068	0.960308	BFGS 13	Identity	Exponential
6	RBF 10-21-1	0.759254	0.718573	0.823115	RBFT	Gaussian	Identity

Table 7: The output of FKHR for different concentration using CNS

S. No	Network Name	Training perfection (%)	Test perfection (%)	Validation perfection (%)	Training algorithm	Hidden activation	Output activation
1	MLP 10-8-1	0.964950	0.960788	0.957878	BFGS 16	Tanh	Identity

CONCLUSION

In this paper we have used the mathematical modeling by linear and non-linear methods to make a best linear model using ten concentrations combination of TNF, EGF and Insulin for FKHR. In this we have find all the results for r^2, r^2_{adj}, PRESS, q^2_{cv}, Durbin Watson statistics, SER, VIF and *t*- values for our 10 data sets which comes out to be correct. Later we have used different distribution functions to find the pdf. We have plotted the probability function, survival function and hazard function using different distributions for FKHR. We have also performed different tests like KS, AD and chi square on different distribution function. For non- linear modeling we have used neural network method using MLP and RBF. Results with RBF techniques are the best in comparison with RBF.

REFERENCES

Weiss, R.,"Cellular computation and communications using engineered genetic regulatory networks". PhD Thesis, MIT, 2001.

Gaudet Suzanne, Janes Kevin A., Albeck John G., Pace Emily A., Lauffenburger Douglas A, and Sorger Peter K. (2005) A compendium of signals and responses trigerred by prodeath and prosurvival cytokines Manuscript M500158-MCP200.
Janes Kevin A, Albeck John G, Gaudet Suzanne, Sorger Peter K, Lauffenburger Douglas A, Yaffe Michael B. Dec.9, 2005 A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis; Science 310**,** 1646-1653.
Brockhaus M, Schoenfeld HJ, Schlaeger EJ, Hunziker W, Lesslauer W, and Loetscher H (1990) Identification of two types of tumor necrosis factor receptors on human cell lines by monoclonal antibodies. Proc Natl Acad Sci USA 87, 3127-3131.
Thoma B, Grell M, Pfizenmaier K, and Scheurich P (1990) Identification of a 60-kD tumor necrosis factor (TNF) receptor as the major signal transducing component in TNF responses. J Exp Med 172, 1019-23.
Libermann TA , Razon TA., Bartal AD, Yarden Y., Schlessinger J and Soreq H 1984 Expression of epidermal growth factor receptors in human brain tumors Cancer Res. 44,753-760.
Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR,, Carotenuto A, De Feo G, Caponiqro F, Salomon DS. 2006 Epidermal growth factor receptor (EGFR) signaling in cancer Gene 366, 2–16.
Ullrich A., Schlessinger J., "Signal transduction by receptors with tyrosine kinase activity", Cell, vol 61, 203-211, 1990.
Arteaga C., "Targeting HER1/EGFR: a molecular approach to cancer therapy" Semin Oncol, vol 30, pp 314, 2003.
Lizcano J. M. Alessi D. R. 2002 The insulin signalling pathway. Curr Biol. 12, 236-238.
Morris F. White 1997 The insulin signaling system and the IRS proteins Diabetologia 40, S2–S17
Morris F. White 2003 Insulin Signaling in Health and Disease Science 302, 1710–1711.
Jain S, Naik P.K., Bhooshan S.V., "Mathematical modeling deciphering balance between cell survival and cell death using insulin", Network Biology, 1(1):46-58, 2011
Jain S, Naik P.K., Bhooshan S.V., A System Model for Cell Death/ Survival using SPICE and Ladder Logic, Digest Journal of Nanomaterials and Biostructures (DJNB), 5(1) : 57-66, 2010, (ISSN 1842 – 3582).
Jain S, Naik P.K. , Sharma R, A Computational Modeling of cell survival/ death using VHDL and MATLAB Simulator, Digest Journal of Nanomaterials and Biostructures (DJNB), 4 (4): 863- 879, 2009, (ISSN 1842 – 3582).
Jain S., "Communication of signals and responses leading to cell survival / cell death using Engineered Regulatory Networks". PhD Thesis, Jaypee University of Information Technology, Solan, Himachal Pradesh, India, 2012.
Mandic, D., Chambers, J., "Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability", John Wiley & Sons, New York, 2001.
Bishop C.M., "Neural Networks for Pattern Recognition", Oxford University Press Oxford, UK, 1995.