Informatics in Medicine Unlocked 23 (2021) 100511

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

Adaptive diagnostic machine learning technique for classification of cell decisions for AKT protein

Ayodeji Olalekan Salau [a,*], Shruti Jain [b]

[a] Department of Electrical/Electronics and Computer Engineering, Afe Babalola University Ado-Ekiti, Nigeria
[b] Department of Electronics and Communications Engineering, Jaypee University of Information Technology, Solan, Himachal Pradesh, India

ARTICLE INFO

Keywords:
AKT Protein
Distribution ID plots
Neural networks
MLP
RBF

ABSTRACT

Artificial intelligence techniques can unravel clinically consistent information in clinical data which in turn assist in decision-making. With the help of state-of-the-art techniques, diagnostic systems are used to identify various diseases by using different machine learning (ML) techniques. In this paper, a combination of ML techniques such as Radial Basis Function (RBF) and Multiple Layer Perceptron (MLP) was used to predict cell decisions (cell survival/death) of AKT protein. AKT signalling networks have various downstream consequences on cellular metabolism either directly through the regulation of nutrient transporters, metabolic enzymes or indirectly through the control of transcription factors that regulate the expression of metabolic pathways which determine cell survival, cell growth, and cell death. Experimental analysis was performed in this work to examine the signalling networks that determine cell survival/death decisions by using an amalgamation of three proteins for ten different combinations in 13 different slices for a period of 0–24 h. *P-P* plot, *Q-Q* plot, and histogram tests were used for data visualization to determine which distribution the data fits. In addition, goodness of fit test was also employed using distribution functions such as Weibull, Exponential, and Normal distribution to determine whether the data fits a distribution of a certain population. The results were validated by calculating the MTTF values. The results of the analysis performed show that the Weibull distribution yields remarkable results. Also the results obtained with the Multiple Layer Perceptron, MLP 10-8-1 was found to perform better than other techniques giving an accuracy of 99.33% when the exponential activation function was used. The results of the experimental study indicate that it is possible to create self-consistent cell-signalling compendia based on AKT protein data that have been computationally simulated to provide valuable insights for cell survival/death regulation.

1. Introduction

Computational biology is a branch of science that involves the application of computer science in the modeling of processes and structures of life [1,2]. It requires the use of computational methods for the simulation of biological systems. Computational biology is used in the modeling of biological systems which help in the sequencing of human genome, and in the modeling of the human brain. In recent times, computational intelligent techniques are used for computational modeling [3]. Computational modeling is used to model real world problems with a view to develop a solution. These models contain different variables that divide the systems understudied. Simulation of these models can be achieved by regulating each of the variables individually or in amalgamation, and by observing how these changes affect their outcome [4]. There are different types of computational modeling techniques which differ from each other in a few dimensions. Non-deterministic, deterministic, static, dynamic, discrete, continuous, stochastic, individual based, popularity logic, and automata are some of the various computational models [5]. Computational modeling is one of the important aspects of big data analytics and nowadays has many applications in healthcare especially for the early diagnosis of diseases [6].

Accurate prediction and diagnosis of multi module and complex systems by making presumptions on the results of various measurements and tests is a general problem occurring in practice [7]. Numerous examples include speech recognition, medical diagnosis, and error-correcting coding. Achieving high diagnostic accuracy requires performing a large number of tests, which can be quite expensive. It is

* Corresponding author.
E-mail addresses: ayodejisalau98@gmail.com (A.O. Salau), shruti.jain@juit.ac.in (S. Jain).

https://doi.org/10.1016/j.imu.2021.100511
Received 20 November 2020; Received in revised form 26 December 2020; Accepted 30 December 2020
Available online 7 January 2021

therefore essential to improve scalability and cost-efficiency of diagnosis by using only the most relevant measurements at any point in time to the current system state and observations. Adaptive learning has its origin in artificialintelligence and started gaining popularity in the 1970s. Artificial intelligence and machine learning technologies have the potential to alter healthcare by deriving new and important insights from the vast amount of data generated during the delivery of health care every day.

Data related to healthcare are disreputable for being complex and voluminous. Computational intelligence is a branch of Artificial intelligence (AI) which aims to transform healthcare by increasing rapid progress of data analytic techniques. Two main components categorize AI methods. The first component consists of machine learning (ML) techniques that analyse organized data such as imaging and genetic data. In this respect, ML is used to gather patient attributes for therapeutic applications or to deduce the probability of disease outcomes. The second portion consists of methods of natural language processing (NLP) that derive information from unstructured data such as medical journals/clinical notes. The NLP procedures target structured data that can be studied using ML techniques that are machine-readable. Academics and medical practitioners believe that applying AI in many practical areas of healthcare will help doctors make better clinical decisions or even substitute human judgment. In 2015, the funds used for the analysis of healthcare data and disease prediction was estimated at $1.48 billion and this is estimated to increase at a rate of 29.3% (annual growth rate) by 2025 [8]. Likewise, the global revenue of $811 million is predicted to rise by 40% by 2021 due to the demand for AI in healthcare applications. The machine learning-as-a-service-market (MLaaS) is projected to hit $5.4 billion by 2022 for the healthcare sector owing to the use of AI [9]. A study of the application of AI in healthcare using artificial neural networks (ANN) to predict if a patient has cardiology was presented in [10]. In healthcare, ANN is used in areas such as speech recognition, clinical diagnosis, prediction of length of stay, cancer prediction [11], medical image analysis [12], and drug development [13]. Improvement in healthcare organizational management requires non-clinical applications [14]. AI aims to impersonate human cognitive functions. With the use of AI in both structured and unstructured healthcare data, AI has brought a paradigm shift in disease diagnosis, detection, and treatment. The ML techniques which are commonly used include neural network [15,16], support vector machine (SVM), decision tree (DT), Naïve Bayes (NB), hidden markov, K-nearest neighbor, logistic regression, discriminant analysis, and deep learning (DL). These techniques have been used in a number of studies to solve healthcare related problems. Authors in Ref. [17] presented a comparative study of SVM technique alongside other ML techniques such as particle swarm optimization (PSO), Quantum particle swarm optimization (QPSO), active set strategy and least square support vector machine (LSSVM) for the diagnosis of cancer. In Ref. [18], a review of ML techniques employed for the diagnosis of neurological disorder was presented, while in Ref. [19], the authors presented a SVM and tongue image approach for the diagnosis of diabetes. In Ref. [20], NB technique was used for the prediction of breast cancer in the United States (US), while authors in Ref. [21] achieved better results using ML techniques such as SVM, NB, K-nearest neighbor, and DT. In Ref. [22], an automatic disease diagnostic system which used SVM was developed, while in Ref. [23], a two dimensional tensor empirical wavelet transform method was used to detect glaucoma eye fundus from normal eye images.

ML is a type of AI technique that encompasses algorithmic techniques which allows machines without specific computer programming to solve problems. The data used by ML techniques must be correctly pre-processed and identified as useful before they can be trained using ML techniques. In ML model building, this is important because the model's accuracy is highly dependent on the reliability of the data in terms of clinical veracity. An efficient ML based disease diagnostic system capitalizes on the doctor's reasoning ability and computer's computing competence. ML techniques are known to provide promising results in healthcare and related areas such as: image registration, multimodal image fusion, computer-aided diagnosis, image annotation, image-guided medical aid, image database retrieval, and medical image segmentation. In the treatment of cancer, there are different types of proteins which leads to cell survival/death which can be identified using different ML techniques. For the diagnosis of cell death/survival in recent years, protein kinase (AKT protein) has been considered as a new area of interest [24,25]. These kinases are used by clinicians for the treatment of chronic inflammatory diseases, cancer, etc. Cancer progression results from complex pathological or physiological processes. The growing masses within the normal tissue or ability to break tissues are the fundamental characteristics of dissemination. This type of spread leads to mortality caused by cancer and also morbidity. The cells which occupy the basement membrane or adjacent healthy tissue for dissemination lead to Metastasis, which is also known as tumor invasion [26]. The processes of invasion includes penetration of surrounding matrix and de-adhesion. The difference between normal invasive growth and abnormal invasion is that normal cells end up with polarized structures, while tumor cells infiltrate into the surrounding tissues. Genomic DNA fragmentation, chromatin condensation, cell shrinkage, membrane blebbing, and disassembly into membrane-bound vesicles (apoptotic bodies) are some of the major apoptotic events. Necrosis and Apoptosis are the various types of cell death [27]. Necrosis is an early cell membrane disruption associated with organelle swelling, while the energy needed for intracellular interaction is activated by Apoptosis, which is tightly regulated and preserved throughout evolution. Apoptotic cell death is known as the progressive series of morphological and biochemical changes in phosphatidylserine cell surfaces, nuclear condensation and fragmentation, proteolytic cleavage of intracellular proteins, and DNA cleavage into nucleosomal fragments. Extrinsic and intrinsic pathways are related with the inspiration of cell death signal pathways in cancer cells [28]. Cytokines, growth factors, and hormones are some extracellular signals which are mediated by receptors that help in the transduction of cellular cues in intracellular physiology [29].

The clear recognition that defects in AKT signalling underlie a diverse array of human diseases makes the clarification of downstream targets and functions of central importance to the understanding and treatment of such diseases. As the ability to knockout, knockdown, or pharmacologically inhibit specific AKT isoforms and related AGC kinases has significantly improved, previously identified substrates should be more rigorously confirmed. Poor regulation of the AKT pathways lead to diseases such as cancer, insulin resistance, diabetes, cardiovascular diseases, and autoimmune diseases. It is regulated by growth signals/factors such as insulin. The AKT pathway leads to glucose uptake and utilization, glycogen synthesis, fatty acid synthesis, protein synthesis, and cell survival/cell proliferation by increasing anabolism and decreasing catabolism.

Due to the analysis of three input proteins, this paper examines the pathways of AKT protein that regulate cell survival/death decisions: tumor necrosis factor-alpha (TNF-alpha) acts as a programmed cell death cue [30,31], while epidermal growth factor (EGF) [32,33] and insulin [34] are markers of survival. The most commonly used deregulated pathway is AKT [35]. Protein Kinase B (PKB), also known as AKT is a threonine/serine protein kinase that plays a key role in various cellular processes such as cell proliferation, glucose metabolism, transcription, cell migration, and apoptosis [36,37]. There are three isoforms of AKT: AKT 1/α, AKT 2/β, and AKT 3/γ as presented in Table 1. These three

Table 1
Types of AKT.

Homology	75–84%	90–95%	73–79%	Chromosomes
	PH	Catalytic Domain	RD	
AKT 1/α		Thr 308	Ser 473	14q32
AKT 2/β		Thr 309	Ser 474	19q13
AKT 3/γ		Thr 305	Ser 472	1q14

Regulatory domain (RD); Pleckstrin homology (PH)

genes are located in different chromosomes of mammals. AKT1 and AKT2 are ubiquitously expressed and AKT3 is found in the kidney, brain, and heart. In Table 1, the domain of Pleckstrin homology (PH) functions as a triphosphate binding module of phosphatidylinositol 3, 4, and 5. The regulatory domain (RD) is located adjacent to the kinase domain in the C-terminus. Phosphoinositide-dependent protein kinase 1 (PDK1), which is also recruited into the plasma membrane by PIP3 is phosphorylated by the 'Thr' residues. While "Ser" residues are targeted by the rapamycin complex 2 (mTOR C2) mammalian target. AKT is inactivated by protein phosphatises action and is first characterised for its function in regulating cell proliferation and survival. Oncogenetic is called the over activation of AKT signalling. All AKT isoforms can phosphorylate the proline rich AKT substrate but the actin-associated protein paladin can only phosphorylate AKT1. AKT is an important therapeutic target for the treatment of human diseases because of AKT's essential role in controlling diverse cellular functions.

In this paper, a ML algorithm was used to enhance competent decisions for AKT related healthcare applications. This paper aims to reduce the research gap with regard to developing successful decision support techniques for cancer related medical applications. Machine learning (ML) is a broad multi-disciplinary area that is rooted in data processing, statistical analytics, and algebra knowledge, etc. Machine-learning techniques are influencing healthcare decisions, i.e. ANN can be used to improve the delivery of healthcare at a reduced cost. A computer assisted disease diagnosis system with a high success rate is important to help doctors diagnose diseases correctly in circumstances that cause experienced doctors to hesitate in making a decision. A current exploratory research in this area shows that new experiments seek to achieve improved results with a high rate of success. The issue of classification can be found in various areas such as the outcomes of medical tests that indicate the existence or absence of a disease. Most academics have recently tried to propose new classification approaches to improve the findings of existing processes. All signalling systems are of great interest since they play an important role in curing diseases such as cancer, rheumatism, and arthritis. The cell decision process requires the analysis of the crosstalk mechanisms between death and survival receptors. Cell proliferation is one essential response governed by signal transduction.

This paper therefore presents an adaptive diagnostic technique (MLP and RBF) for the purpose of predicting cell decisions (cell death/survival). The AKT protein pathways that controls cell death/survival decisions using a combination of three input proteins - *tumor necrosis factor-α* (*TNF*), *epidermal growth factor* (*EGF*), and *insulin* was examined. AKT regulates cell metabolism by using binding, cellular survival, and different downstream effectors. Different parameters were calculated for all the ten different combinations of three inputs proteins for AKT protein. Out of which the best combination of TNF-EGF-Insulin was identified using different features like maximum, minimum, standard deviation, and mean values, and the results were validated by calculating the observed, expected, and cumulative values for different boundary levels. The selected concentrations were used for the classification of the cells using ANN techniques such as MLP and RBF. A time series 3D plot was generated for all the best combinations and validated with the training and testing accuracies. The training and testing results yield the same results as that of the ANN model which accurately predicts cell survival or otherwise cell death.

The novelty of this paper lies in the use of *Q-Q* plot and histogram test for data visualization which was used to observe the distribution of data. Goodness of fit test was used to determine whether the data fits a distribution of a certain population and the results were validated using MTTF test. Normality test was also performed using distribution ID plots considering different distribution functions. Different features were extracted and were classified using ANN techniques such as RBF and MLP by varying the activation functions instead of using other ML or deep learning techniques [38].

The remaining paper is structured as: Section 2 presents the materials and methods used for the prediction of cell survival/death for AKT protein. Section 3 describes the findings obtained and the implications are discussed afterward. This is accompanied by a conclusion and suggestion for future work in section 4.

2. Materials and methodology

In this section, experimental analysis was performed on HT29 carcinoma cells to examine the pathways of AKT which help in controlling survival/death decisions due to the amalgamation of three different proteins. The AKT/Protein kinase B (PKB)/Ras-related C3 botulinum toxin substrate (Rac) are the inceptive kinase with indistinguishable properties as Protein kinase C (PKC)/Protein kinase A (PKA). These proteins take part in metabolism, progression of survival and apoptosis. Phosphatidylinositol-3-kinase (PI-3K)/AKT signalling pathway is operated through tyrosine kinase (TrK). The phosphoinositide phosphates (PIP2 and PIP3) are two identical monomer molecules which composed of the plasma membrane and Protein kinase 1 (PDK1). PDK1 combines two identical monomers leading to activation of AKT. AKT leads to several types of cancers. It regulates the metabolism by binding, cellular survival, and different downstream effectors. For transcription of pro-survival genes, AKT activates NFkB via IKB kinase. In this research paper, the work is divided into two modules: experimental analysis and computational modeling. The computational model for cell survival/death was developed using an algorithm which was implemented using Statistica Software 11 [39] and Minitab 11 software on an Intel Core processor system. This algorithm is presented in Table 2. In total, 300 values of each combination of input proteins was considered in this paper.

2.1. Experimental analysis

For the experimental analysis, authors have considered HT29 carcinoma cells. The human colon adenocarcinoma cell line HT29, is not only used to study the biology of human colon cancers, but is receiving special interest in studies focused on food digestion and bioavailability due to the ability to express characteristics of mature intestinal cells. The permanent cell line HT29 was established by Fogh and Trempe in 1975 [40] from an adenocarcinoma of human colon. HT29 carcinoma cells are sensitive to different drugs such as oxaliplatin, 5-fluorouracil, etc. These cells are also sensitive to chemotherapeutic drugs that are used for the treatment of colorectal cancer. The HT29 cell line is also employed as an *in-vitro* model to learn transportation, absorption, and secretion of intestinal cells. The experimental analysis was performed to examine the different combinations of three proteins (0, 0.2, 1, 5, 100, 500 ng/ml) for a 24 h time period. The 24 h' time is partitioned as 0, 5, 15, 30, 60,

Table 2
Algorithm of the computational modeling procedure for AKT protein.

Input: HT carcinoma cells
Output: Computational model resulting in the determination of cell survival/death
Step 1: Perform experimental analysis on HT carcinoma cells considering three different proteins.
Step 2: Perform analysis for different concentration values (ng/ml) and at different time periods.
Step 3: Prepare heat map for the resultant considering only AKT protein.
Step 4: Pre-processing the resultant data using goodness of fit test considering different distribution functions such as Weibull, Exponential, and Normal distribution.
Step 5: Validate the results by calculating MTTF values for different distribution functions.
Step 6: Extract the various features for testing, training, validation, and overall phases considering the ten different concentrations.
Step 7: Perform classification using the ANN Machine learning technique (MLP and RBF).
Step 8: Computational model output results in the prediction of cell survival/death to diagnose cancer and AKT related diseases.

and 90 min, and 2, 4, 8, 12, 16, 20, and 24 h. To investigate the efficient relationships between the cytokine-receptor interactions, commencement of intracellular signalling cascades, the survival-death decisions were interpreted by analysing ten disparate treatments of input proteins. The analysis was done for thirteen different proteins, but in this paper authors only examined the AKT signals. This study generates a heat map of the three different proteins (TNF-EGF-Insulin) with different concentrations (0-0-0, 5-0-0, 5-1-0, 0-0-500, 100-0-0, 5-0-5, 0-100-0, 100-100-0, 0.2-0-1, and 100-0-500).

2.2. Computational modeling

In this research paper, *P-P* plot, *Q-Q* plot, and histogram tests were used for data visualization so as to observe the distribution of data. Goodness of fit test was also employed considering distribution functions such as Weibull, Exponential, and Normal distribution to determine whether the data fits a distribution of a certain population. The results were validated by calculating MTTF values. After the fiting and normalization of data, different features were extracted. Feature extraction is the process of reduction of redundant attributes of an image and simplifying the amount of resources required to represent an image [41]. The temporal features (time domain features), which are simple to extract and have easy physical interpretation are: the energy of the signal, zero crossing rate, maximum amplitude, minimum energy, etc. The spectral features (frequency based features), which are obtained by converting the time based signal into the frequency domain using Fourier Transform, like: fundamental frequency, frequency components, spectral centroid, spectral flux, spectral density, and spectral roll-off. Various features such as minimum, maximum, mean, and standard deviation were calculated for testing, validation, training, and overall data for all ten different concentrations of the three input combinations. The extracted features were classified using the proposed ML technique (MLP and RBF). The proposed methodology is shown in Fig. 1. Machine learning offers the capability of efficient and effective classification of different sensed imagery. The power of ML provides the ability to map classes with very multifaceted characteristics and to handle data of high dimensionality. In these modern times, the accessibility of large datasets coalesced with the exponential growth in calculating power and enhancement in algorithms led to an unparalleled flow of awareness in the area of ML. Currently, ML algorithms are frequently used for dimensionality reduction, clustering, regression, or classification of high-dimensional input data. Nevertheless, executing a ML classification

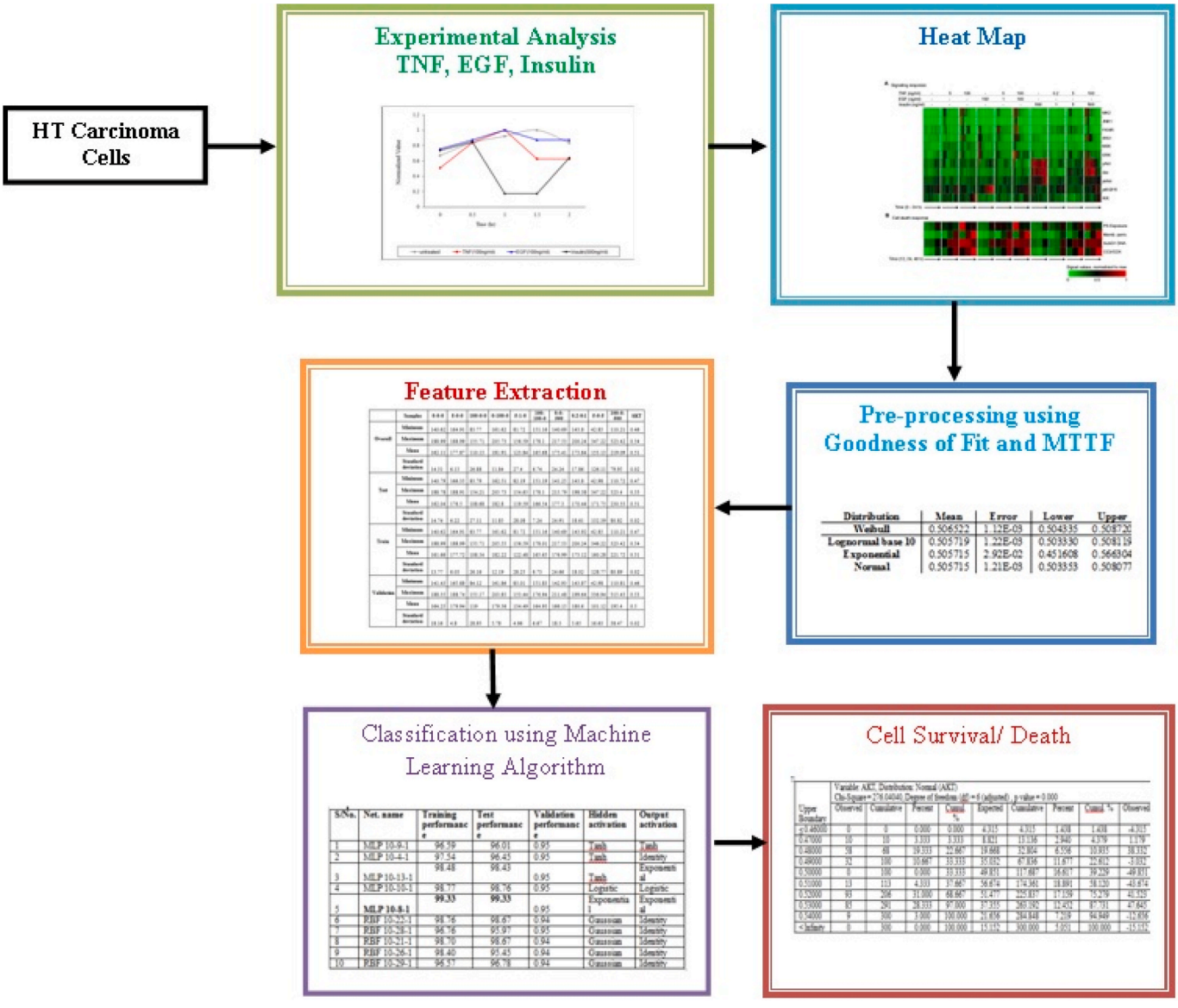

Fig. 1. Proposed methodology for cell death/survival decision using AKT protein.

is not simple and recent literatures provide divergent recommendations regarding many concerns. Typically, all ML algorithms are categorized into groups based on function, learning style, or the problem to be solved. In this paper, artificial neural network (ANN) is used for classification purpose. In engineering, ANN can be used for two important functions, namely as a pattern classifier and nonlinear adaptive filter. They are mainly adaptive nonlinear systems that study the performance of an input/output map function of a dataset. Adaptive means that during the process, the device parameters adjust and this is typically referred to as the training period. The ANN parameters are permanent after the training process and can be used to solve problems. The ANN method is roughly categorized into MLP and RBF. Between inputs, weights and sigmoidal activation functions, MLP uses dot products. By using back propagation for all layers, training is typically accomplished. RBF utilizes Euclidean weight and input distances which can be used as Gaussian activation functions that make neurons more locally sensitive. RBF possibly uses hybrid approaches with unsupervised learning or back propagation for learning.

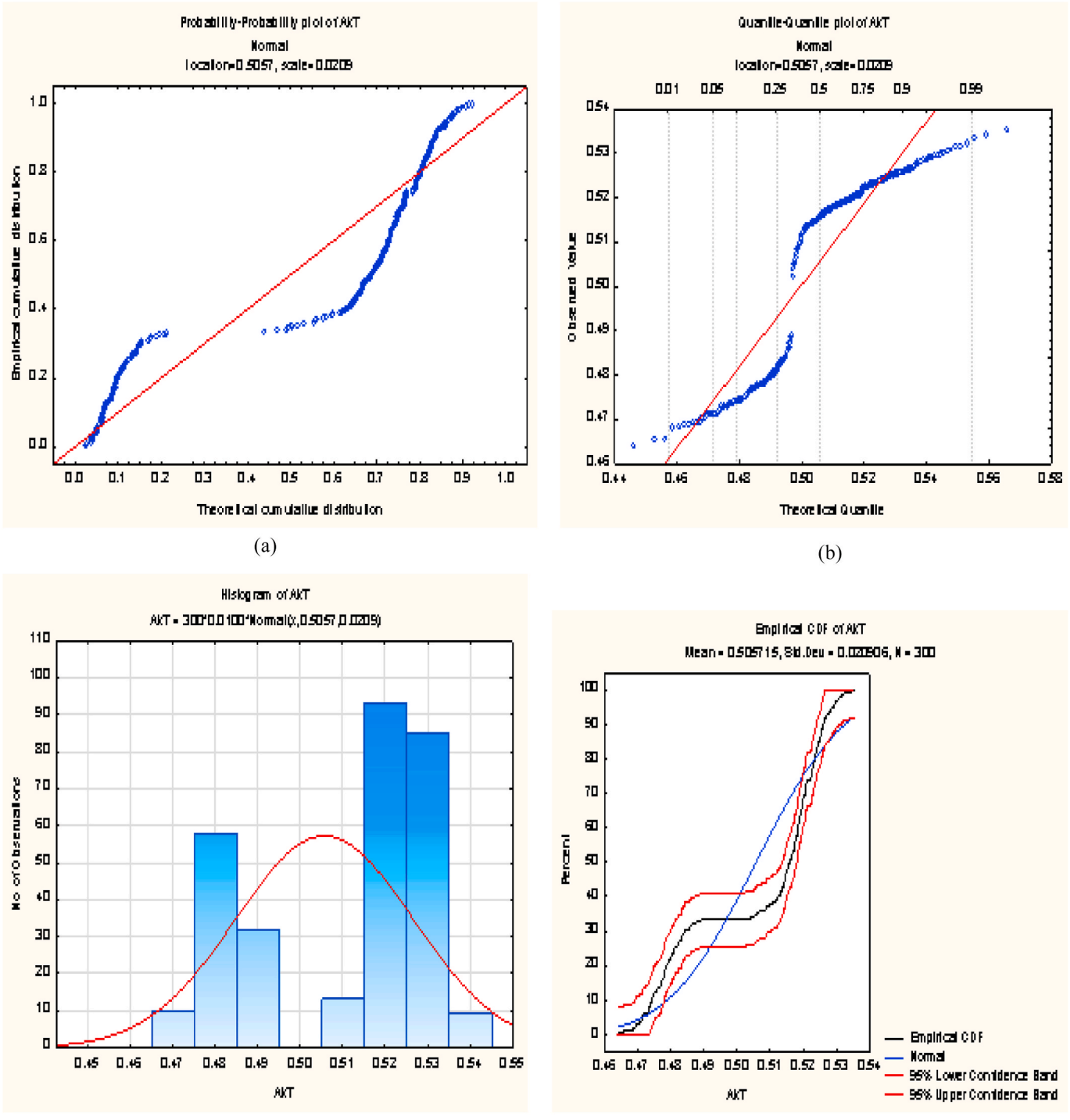

Fig. 2. Preprocessing for AKT.

3. Results and discussion

As discussed in the methodology section regarding experimental analysis, authors have performed experiments and the results of the experimental analysis is presented and discussed in this section. The results of the experimental analysis showed the uniqueness of the three-fold marker protein selection technique; the first stage which included pre-processing techniques, followed by extraction of different features like maximum, minimum, standard deviation, and mean values to select the best combinations of TNF-EGF-Insulin, and lastly, the detection stage which was performed using MLP and RBF technique to provide a high prediction accuracy and low complexity. The results obtained when the input proteins response of HT29 human colon carcinoma cells was examined indicate that the MLP and RBF approach can be used to reveal important aspects of biological cue-signal-response systems. In genomics, the databases used for these types of experiments are prevalent, primarily because they contain sequence data with consistent start and end points with ease of fusing data which are organized and homogeneous. Cell-signalling data on the other hand is unstructured and heterogeneous and relies on biological content. Ten different concentrations were measured and then normalized with three inputs and an average of four outputs (phosphatidylserine exposure (PE), membrane permeability (MP), nuclear fragmentation (NF) and caspase substrate cleavage (CC)), giving a final result of 10 inputs and 1 output.

The results of the generated heat map are shown in Fig. 2 in form of images pre-processed by plotting histograms, *P-P* plot, *Q-Q* plot, and performing normality tests. Fig. 2(a) shows the *P-P* plot, Fig. 2(b) represents the *Q-Q* plot, Fig. 2(c) gives the histogram representation, and Fig. 2(d) represents the empirical cumulative distribution function (CDF) of AKT. The goodness of fit using Anderson Darling (AD) values were calculated for different distribution functions and the results are presented in Table 3. The goodness of fit test is a statistical test that helps to determine whether a set of observed values match those expected relative to the understudied model. There are different types of goodness of fit tests, but in this paper, AD test was used. This test helps to determine whether the data fits into an expected set of data. This is also known as test of normality. This test doesn't make any assumption about the distribution of the data. AD statistic is used to compare the degree to which a set of data fits into a certain distribution function to determine which one is best. The best function is obtained when the AD value is lower than others [42]. Table 3 shows that the Weibull distribution function yields remarkable results. To validate the results, the percentile and mean time to failure (MTTF) tests were also performed. Most times it is very difficult to determine the best distribution on the basis of probability plot and goodness of fit measures. The table of percentiles (shown in Table 4) helps in comparing the percentiles for various different distribution functions. The table shows that based on Normal distribution fitted to cell death, 1% of the cells are expected to fail by 0.45, likewise 10% of the cells are expected to fail by 0.47. MTTF was used to compare the selected distribution with a distribution ID plot. The normal function has a MTTF of 0.505 as tabulated in Table 5. The Weibull function has the largest MTTF of 0.506. The Weibull function is selected because of its better MTTF value which validates our results. The exponential mean or Mean Time to Fail (MTTF) is calculated using $1/\lambda$, where λ is a constant for the time which reduces with failure rate. The exponential distribution is the only distribution to have a constant failure rate. The different features extracted from the ten different combinations of three input proteins for AKT protein are presented in Table 6. For all the ten combinations of protein, 42.85 was obtained as the minimum value, 347.22 as the maximum value, 153.13 as the mean value, and 126.11 as the standard deviation for 5-0-5 ng/ml combinations of input proteins. Where 5-0-5 ng/ml signifies 5 ng/ml of TNF, 0 ng/ml of EGF, and 0 ng/ml of insulin.

The selected concentrations were used to classify the cells using the MLP and RBF ANN techniques. If the ANN techniques expected performance is > 0.5, it will lead to cell survival, or else it will lead to cell death. For MLP 10-8-1, 10 means ten distinct combinations of three distinct input proteins, 8 hidden layers, and 1 output prediction of cell survival/death. The MLP 10-8-1 3D plot for the final result is shown in Fig. 3. Different threshold values expressed with different colors are used to distinguish between cell survival/death. The accuracies of the training and testing accuracies of ten different ANN models using MLP and RBF are presented in Table 7. The table shows that MLP 10-8-1 gives 99.33% training accuracy and 99.33% of testing accuracy. It was also observed that the maximum individual classification accuracy for survival is 98.66% and maximum individual classification accuracy for death is 100%. Out of 300 testing cases, 148 cases (148/150) were

Table 3
Goodness of fit test for AKT protein.

Distribution	Anderson-Darling (adj)
Weibull	18.44
Lognormal base 10	23.28
Exponential	126.69
Normal	22.63

Table 4
Percentiles for AKT protein.

Distribution	Percentage	Percentile	Error	Lower	Upper
Weibull	1	0.447927	3.43E-03	0.441256	0.454698
Lognormal base 10	1	0.458430	2.13E-03	0.454272	0.462626
Exponential	1	0.005083	2.93E-04	0.004539	0.005692
Normal	1	0.457161	2.32E-03	0.452614	0.461708
Weibull	5	0.470660	2.50E-03	0.465793	0.475579
Lognormal base 10	5	0.471685	1.75E-03	0.468273	0.475121
Exponential	5	0.025940	1.50E-03	0.023164	0.029048
Normal	5	0.471385	1.85E-03	0.467762	0.475007
Weibull	10	0.481064	2.07E-03	0.477027	0.485136
Lognormal base 10	10	0.478907	1.56E-03	0.475858	0.481975
Exponential	10	0.053282	3.08E-03	0.047582	0.059666
Normal	10	0.478967	1.63E-03	0.475780	0.482154
Weibull	50	0.509393	1.05E-03	0.507331	0.511464
Lognormal base 10	50	0.505277	1.22E-03	0.502891	0.507674
Exponential	50	0.350535	2.02E-02	0.313031	0.392532
Normal	50	0.505715	1.21E-03	0.503353	0.508077

Table 5
Values of MTTF for AKT protein.

Distribution	Mean	Error	Lower boundary	Upper boundary
Weibull	0.506522	1.12E-03	0.504335	0.508720
Lognormal base 10	0.505719	1.22E-03	0.503330	0.508119
Exponential	0.505715	2.92E-02	0.451608	0.566304

Table 6
Different extracted features for AKT protein.

	Samples	0-0-0	5-0-0	100-0-0	0-100-0	5-1-0	100-100-0	0-0-500	0.2-0-1	5-0-5	100-0-500	AKT
Overall	**Minimum**	140.62	164.91	83.77	161.62	81.72	151.16	140.69	143.8	42.85	110.21	0.46
	Maximum	188.99	188.99	155.71	205.73	156.59	178.1	217.53	200.24	347.22	323.42	0.54
	Mean	162.11	177.87	110.13	181.91	123.84	165.68	175.41	173.84	153.13	219.09	0.51
	Standard deviation	14.31	6.15	26.88	11.84	27.4	6.74	24.24	17.86	126.11	79.95	0.02
Test	**Minimum**	140.79	166.35	83.79	162.51	82.19	151.19	141.25	143.8	42.98	110.72	0.47
	Maximum	188.78	188.91	154.21	205.73	154.65	178.1	215.79	198.38	347.22	323.4	0.53
	Mean	162.04	176.5	108.68	182.8	119.59	166.54	177.3	170.44	171.75	230.53	0.51
	Standard deviation	14.74	6.22	27.11	11.85	28.08	7.24	24.91	18.61	132.39	80.82	0.02
Train	**Minimum**	140.62	164.91	83.77	161.62	81.72	151.16	140.69	143.92	42.85	110.21	0.47
	Maximum	188.99	188.99	155.71	205.55	156.59	178.01	217.53	200.24	346.22	323.42	0.54
	Mean	161.66	177.72	108.54	182.22	122.48	165.65	176.99	173.12	160.28	221.72	0.51
	Standard deviation	13.77	6.03	26.16	12.19	28.25	6.73	24.66	18.02	128.77	80.89	0.02
Validation	**Minimum**	141.43	165.88	84.12	161.86	83.01	151.83	142.93	143.87	42.98	110.81	0.46
	Maximum	188.35	188.74	155.17	203.85	153.44	176.94	211.48	199.64	336.94	315.43	0.53
	Mean	164.25	179.94	119	179.56	134.49	164.93	166.15	180.6	101.12	195.4	0.5
	Standard deviation	18.16	4.8	28.95	5.78	4.96	6.67	18.3	5.65	16.63	38.47	0.02

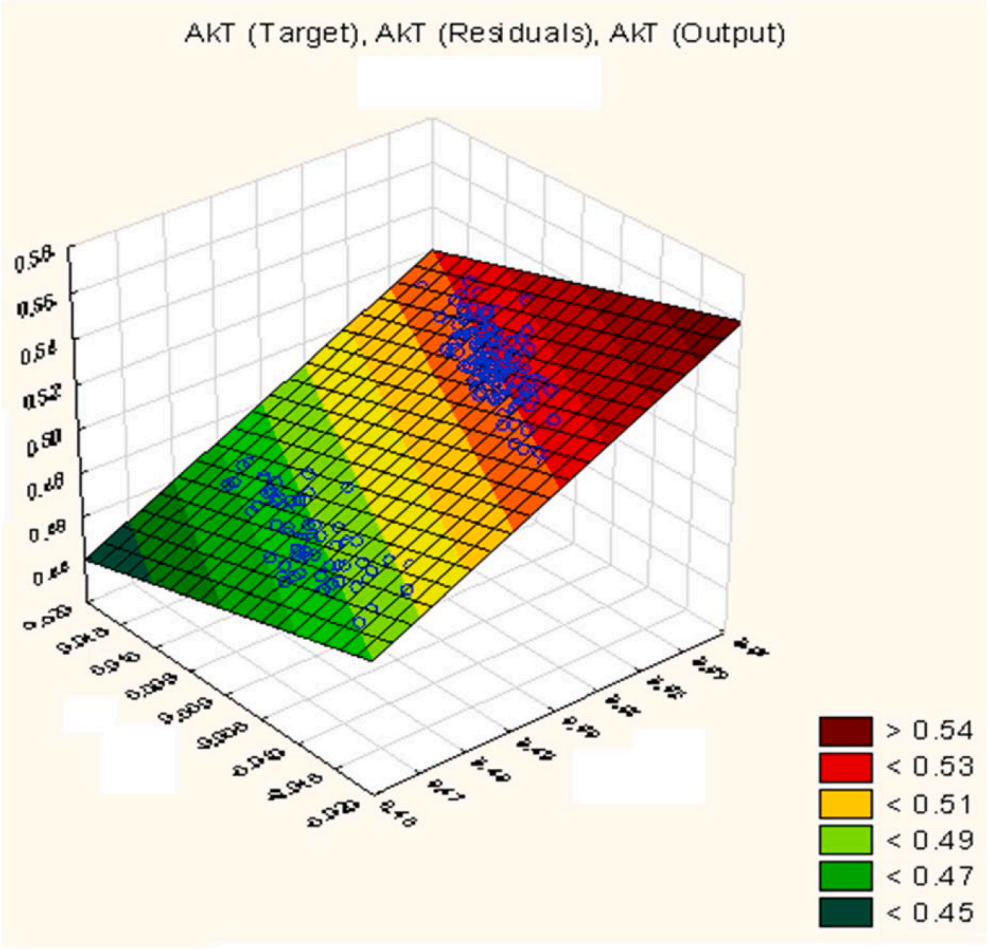

Fig. 3. 3D plot of MLP 10-8-1 for Final output of residual, target, and output.
Note: 10 signifies ten different combinations of three different input protein, 8 hidden layers, 1 output cell survival/death.

correctly classified and 3 cases (3/150) were wrongly classified. Finally, it was observed that the training and testing results yield the same value as that of the ANN model (MLP 10-8-1), indicating the accurate prediction of cell survival or else death. The designed algorithm when tested on AKT protein shows that the MLP provides better results with the least run-time complexity for cell survival/death prediction. Since, ANNs techniques are adaptive to complex problems, changing the networks topology makes them able to handle different levels of complexity and predict the desired output of a system when adequate experimental data is provided. One of the advantages of neural network is that it allows the modeling of physical phenomena in complex systems without requiring exhaustive experimentation or without requiring explicit mathematical representations. Non–linear neural network was used to uncover important aspects of biological cue-signal-response systems using TNF, EGF, and Insulin mediated response of HT-29 human colon carcinoma cells. Although several analysis were performed, the hallmark of this work is in the description of the predictive model of a cytokine-signal-response compendium used to investigate the regulation

Table 7
Comparison of training and testing accuracies using different ANN models for AKT.

S/No.	Network name	Training performance	Test performance	Validation performance	Hidden activation	Output activation
1	MLP 10-9-1	96.59	96.01	0.95	Tanh	Tanh
2	MLP 10-4-1	97.54	96.45	0.95	Tanh	Identity
3	MLP 10-13-1	98.48	98.43	0.95	Tanh	Exponential
4	MLP 10-10-1	98.77	98.76	0.95	Logistic	Logistic
5	**MLP 10-8-1**	**99.33**	**99.33**	0.95	Exponential	Exponential
6	RBF 10-22-1	98.76	98.67	0.94	Gaussian	Identity
7	RBF 10-28-1	96.76	95.97	0.95	Gaussian	Identity
8	RBF 10-21-1	98.70	98.67	0.94	Gaussian	Identity
9	RBF 10-26-1	98.40	95.45	0.94	Gaussian	Identity
10	RBF 10-29-1	96.57	96.78	0.94	Gaussian	Identity

of cell fate with the combination of the input proteins for AKT protein. The compendium contains more than 10,000 biochemical measurements based on the states and activities of cell-signalling proteins and apoptotic responses in human cells. Experimental databases are common in genomics, majorly because sequence data are structured and homogeneous with clear start and finish points, and they provide a relative ease to fuse data. In contrast, cell-signalling data are unstructured and heterogeneous, and depend on biological content. The results were validated by calculating the observed, expected, and cumulative values for different boundaries level as tabulated in Table 8.

In Table 8, the observed frequency (f_{obs} or f_{exp}) is the number of individual data points in a specific class. To calculate the total number of individual data points in a dataset, sum all the frequencies resulting to *N*. The percent observed is the ratio of f_{obs} and *N,* (*fobs/N*). The cumulative frequency (cf) represents the number of individual data data points located at or below each score and its percentages represent the percentage of data accumulated as it moves up the scale.

A comparison of existing works is also presented in Table 9. Experimental findings show that the proposed technique performs better than existing techniques, with the ability to provide low complexity and high prediction accuracy. In addition, it showed that self-consistent cell signalling data based on AKT protein which are replicated computationally can give insights into the control of cell decisions.

4. Conclusion

Health related problems in the medical field are enormous and treatment is becoming more expensive. In recent times, machine learning (ML) techniques are being used as a vital tool to solve the challenges in this area. This study showed that ML techniques such as Radial Based Function (RBF) and Multiple Layer Perceptron (MLP) can be used for different levels of healthcare related decision-making and diagnosis to aid treatment. Influenced by the ever increasing advancements in the field, medical experts take advantage of the hybrid models of ANN with efforts to tailor solutions to healthcare related problems. This paper presented a series of experimental analysis to predict cell decisions of AKT protein using ten different combinations of three input proteins over a span of 0–24 h in 13 different slices. The mechanisms for activating AKT/PKB, the lipid second messenger-mediated phosphorylation of AKT/PKB have been well described in recent decades, and the study of AKT/PKB in promoting cell survival and proliferation is rapidly expanding. The data was visualised and normalized for the purpose of experimentation to extract various features. Lastly, RBF and MLP techniques were applied for the prediction of cell death/cell survival. Promising results were achieved using RBF and MLP techniques for the prediction of cell survival/death. In light of this, we hope that a critical appreciation of the way they operate and the risks they might pose can serve as a motivation for academics and medical specialist to develop new and transformative ML techniques to better diagnose cancer and its related diseases. In the future, the authors hope to explore the use of deep learning to achieve a better performance for cell survival, which in turn will entail a variety of optimization techniques to be applied for intensified selection of attributes.

Table 9
Comparison of proposed technique with related works.

S/No.	Technique	Proteins	Accuracy (%)
1.	GLCM + k-NN [32]	EGFR, IRS, ERK, MK2, JNK, FKHR	75.60
2.	GLDS + k-NN [28]	AKT	76.90
3.	DWT + SSVM [36]	ERK, MK2, JNK	80.00
4.	GLDS + SVM [28]	AKT	84.60
5.	GLCM + SVM [32]	EGFR, IRS, ERK, MK2, JNK, FKHR	85.80
6.	Proposed Technique (MLP and RBF)	AKT	99.33

Funding

Authors declare no funding for this research.

Table 8
Fitting distribution and chi-square value for different boundaries level.

Upper Boundary	Variable: AKT, Distribution: Normal (AKT) Chi-Square = 276.04040, Degree of freedom (df) = 6 (adjusted), p value = 0.000							
	Observed Frequency (f_{obs})	Percent Observed	Cumulative Frequency (cf)	Cumul. % Observed	Expected Frequency (f_{exp})	Percent Expected	Cumulative Expected (cf)	Cumul. % Expected
≤0.46000	0	0.000	0	0.000	4.315	1.438	4.315	1.438
0.47000	10	3.333	10	3.333	8.821	2.940	13.136	4.379
0.48000	58	19.333	68	22.667	19.668	6.556	32.804	10.935
0.49000	32	10.667	100	33.333	35.032	11.677	67.836	22.612
0.50000	0	0.000	100	33.333	49.851	16.617	117.687	39.229
0.51000	13	4.333	113	37.667	56.674	18.891	174.361	58.120
0.52000	93	31.000	206	68.667	51.477	17.159	225.837	75.279
0.53000	85	28.333	291	97.000	37.355	12.452	263.192	87.731
0.54000	9	3.000	300	100.000	21.656	7.219	284.848	94.949
< Infinity	0	0.000	300	100.000	15.152	5.051	300.000	100.000

Availability of data and material

Not applicable.

Code availability

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors declare that they have acknowledged all sources used and any support given.

References

[1] Zhang X, Tang N, Hadden TJ, Rishi AK. Akt, FoxO and regulation of apoptosis. Biochim Biophys Acta 2011;1813:1978–86.
[2] Jain S. System modeling of AKT using linear and robust regression analysis. Curr Trends Biotechnol Pharm 2018;12(2):177–86.
[3] Shahid AH, Singh MP. Computational intelligence techniques for medical diagnosis and prognosis: problems and current developments. Biocybernet Biomed Eng 2019; 39:638–72. https://doi.org/10.1016/j.bbe.2019.05.010.
[4] Farrell PJ, Rogers-Stewart K. Comprehensive study of tests for normality and symmetry: extending the Spiegel halter test. J Stat Comput Simulat 2006;76(9): 803–16.
[5] Xie L, Li Z, Zhou Y, He Y, Zhu J. Computational diagnostic techniques for electrocardiogram signal analysis. Sensors 2020;20:6318. https://doi.org/10.3390/s20216318.
[6] Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L. An analytical method for diseases prediction using machine learning techniques. Comput Chem Eng 2017;106: 212–23. https://doi.org/10.1016/j.compchemeng.2017.06.011.
[7] Internet of things and big data technologies for next generation healthcareBhatt C, Dey N, Ashour AS, editors. Studies in Big Data 2017. https://doi.org/10.1007/978-3-319-49736-5.
[8] Grand View Research. Healthcare predictive analytics market analysis by application (operations management, financial, population health, clinical), by end-use (payers, providers), by region (north America, europe, asia pacific, Latin America, MEA) and segment forecasts. 2016. p. 2018–25.
[9] Tatcher L. The dangers of commoditized machine learning in healthcare: 5 key differentiators that lead to success: Healthcare. Available from: https://healthcare.ai/dangers-ofcommoditized-machine-learning-in-healthcare/; 2018.
[10] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4):230–43. https://doi.org/10.1136/svn-2017-000101.
[11] Lee CW, Park JA. Assessment of HIV/AIDS-related health performance using an artificial neural network. Inf Manag 2001;38(4):231–8. https://doi.org/10.1016/S0378-7206.
[12] Khare A, Jeon M, Sethi IK, Xu B. Machine learning theory and applications for healthcare. J Healthcare Eng 2017:1–2. https://doi.org/10.1155/2017/5263570.
[13] Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. PloS One 2019;14(2): e0212356. https://doi.org/10.1371/journal.pone.0212356.
[14] Goss EP, Vozikis GS. Improving health care organizational management through neural network learning. Health Care Manag Sci 2002;5(3):221–7.
[15] Kaur H, Wasan SK. Empirical study on applications of data mining techniques in healthcare. J Comput Sci 2006;2(2):194–200.
[16] Nolting J. Developing a neural network model for health care. AMIA Annu Symp Proc 2006;1049.
[17] Sweilam NH, Tharwat AA, Abdel Moniem NK. Support vector machine for diagnosis cancer disease: a comparative study. Egyptian Informatics J 2010;11: 81–92.
[18] Raghavendra U, Acharya UR, Adeli H. Artificial intelligence techniques for automated diagnosis of neurological disorders. Eur Neurol 2019:1–24. https://doi.org/10.1159/000504292.
[19] Zhang J, Xu J, Hu X, Chen Q, Tu L, Huang J, Cui J. Diagnostic method of diabetes based on support vector machine and tongue images. BioMed Res Int 2017:1–9. https://doi.org/10.1155/2017/7961494.
[20] Williams, et al. Breast cancer risk prediction using data mining classification techniques. Trans Networks Commun 2015;2:1–11.
[21] Senturk, et al. Breast cancer diagnosis via data mining: performance analysis of seven different algorithms. Comput Sci Eng 2014;1:1–10.
[22] Gürbüz E, Kılıç E. A new adaptive support vector machine for diagnosis of diseases. Expet Syst 2013;31(5):389–97. https://doi.org/10.1111/exsy.12051.
[23] Jain S, Salau AO. Detection of glaucoma using two dimensional tensor empirical wavelet transform. SN Applied Sci 2019;1(11):1417. https://doi.org/10.1007/s42452-019-1467-3.
[24] Salau AO, Jain S. Computational modeling and experimental analysis for the diagnosis of cell survival/death for AKT protein. J Genetic Eng Biotechnol 2020;18 (11):1–10. https://doi.org/10.1186/s43141-020-00026-w.
[25] Kamal MS, Chowdhury L, Khan MI, Ashour AS, Tavares JMRS, Dey N. Hidden Markov model and Chapman Kolmogrov for protein structures prediction from images. Comput Biol Chem 2017;68:231–44. https://doi.org/10.1016/j.compbiolchem.2017.04.003.
[26] Van der Kamp MW, Schaeffer RD, Jonsson AL, Scouras AD, Simms AM, Toofanny RD, Daggett V. Dynameomics: a comprehensive database of protein dynamics. Structure 2010;18(4):423–35. https://doi.org/10.1016/j.str.2010.01.012.
[27] de Bruin EC, Medema JP. Apoptosis and non-apoptotic deaths in cancer development and treatment response. Canc Treat Rev 2008;34:737–49.
[28] Jain S, Salau AO. An image feature selection approach for dimensionality reduction based on KNN and SVM for AKT proteins. Cogent Eng 2019;6(1):1–14. https://doi.org/10.1080/23311916.2019.1599537.
[29] Kevin JA, John AG, Suzanne G, Peter SK, Douglas LA, Michael YB. A systems model of signalling identifies a molecular basis set for cytokine-induced apoptosis. Science 2005;310:1646–53.
[30] Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, Carotenuto A, De Feo G, Caponiqro F, Salomon DS. Epidermal growth factor receptor (EGFR) signaling. Can Geneal 2006;366:2–16.
[31] Brunet A, Bonni A, Zigmond MJ, Lin MZ, Juo P, Hu LS, Anderson MJ, Arden KC, Blenis J, Greenberg ME. AKT promotes cell survival by phosphorylating and inhibiting a fork head transcription factor. Cell 1999:857–68.
[32] Jain S, Chauhan DS. Instance-based learning of marker proteins of carcinoma cells for cell death/survival. Comput Methods Biomech Biomed Eng: Imag Visual 2019. https://doi.org/10.1080/21681163.2019.1692236.
[33] Kamal MS, Sarowar MG, Dey N, et al. Self-organizing mapping based swarm intelligence for secondary and tertiary proteins classification. Int. J. Mach. Learn. & Cyber. 2019;10:229–52. https://doi.org/10.1007/s13042-017-0710-8.
[34] Gaudet S, Kevin JA, John AG, Emily PA, Douglas LA, Peter SK. A compendium of signals and responses triggered by pro-death and pro-survival cytokines. 2005. Manuscript M500158-MCP200.
[35] Manning BD, Cantley LC. AKT/PKB signaling navigating downstream. Cell 2007; 129:1261–74.
[36] Kim JS, Ku B, Woo TG, Oh AY, Jung YS, Soh YM, Ha NC. Conversion of cell-survival activity of AKT into apoptotic death of cancer cells by two mutations on the BIM BH3 domain. Cell Death Dis 2015;6(7):e1804. https://doi.org/10.1038/cddis.2015.118.
[37] Jain S. Classification of protein kinase b using discrete wavelet transform. Int J Inf Technol 2018;10(2):211–6. https://doi.org/10.1007/s41870-018-0090-7.
[38] Lan K, Wang D, Fong S, Liu L, Wong KKL, Dey N. A survey of data mining and deep learning in bioinformatics. J Med Syst 2018;42(8). https://doi.org/10.1007/s10916-018-1003-9.
[39] STATISTICA Software Link. https://statistica.software.informer.com/.
[40] Fogh J, Trempe G. New human tumor cel lines. In: Fogh J, editor. Human tumor cells in vitro. Boston, MA: Springer; 1975. https://doi.org/10.1007/978-1-4757-1647-4_5.
[41] Salau AO, Jain S. Feature extraction: a survey of the types, techniques, and applications. In: 5th IEEE international conference on signal processing and communication (ICSC), noida, India; 2019. p. 158–64. https://doi.org/10.1109/ICSC45622.2019.8938371.
[42] Wong WK, Sim CH. Goodness-of-fit based on empirical characteristic function. J Stat Comput Simulat 2000;65:243–69.