

The anti-proliferative potential of Albizia lebbeck ethanolic extract on human breast cancer cells and performance of extreme machine models in modelling of cell migration

Huzaifa Umar (✉ Huzafa.umar@neu.edu.tr)

Near East University

Abdullahi Garba Usman

Near East University

Maryam Rabi

Cyprus International University

Umar Muhammad Ghali

Near East University

Sani Isah Abba

King Fahd University of Petroleum & Minerals

Dilber Uzun Ozsahin

University of Sharjah

Article

Keywords: Cell Migration, Cytotoxicity, Multilayer Perceptron, Extreme Gradient Boosting, Extreme Learning Machine

Posted Date: August 10th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3142376/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on December 14th, 2023. See the published version at <https://doi.org/10.1038/s41598-023-49363-z>.

The anti-proliferative potential of *Albizia lebbbeck* ethanolic extract on human breast cancer cells and performance of extreme machine models in modelling of cell migration

Huzaifa Umar ^{1,*}, Abdullahi Garba Usman ^{1,2}, Maryam Rabi'u ³, Umar Muhammad Ghali ⁴, Sani Isah Abba ⁵, Dilber Uzun Ozsahin ^{7,6,1},

¹ Near East University, Operational Research Center in Healthcare, Nicosia 99138, TRNC Mersin 10, Turkey.

² Department of Analytical Chemistry, Faculty of Pharmacy, Near East University, TRNC, Mersin 10, 99138, Nicosia, Turkey.

³ Department of Energy System Engineering Cyprus International University, Nicosia 99258, Northern Cyprus via Mersin 10 Turkey.

⁴ Department of Medical Biochemistry, Faculty of Medicine, Near East University, TRNC, Mersin 10, 99138, Nicosia, Turkey.

⁵ Interdisciplinary Research Center for Membranes and Water Security, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia.

⁶ Department of Medical Diagnostic Imaging, College of Health Sciences, University of Sharjah, Sharjah P.O. Box 27272, United Arab Emirates.

⁷ Research Institute for Medical and Health Sciences, University of Sharjah, Sharjah P.O. Box 27272, United Arab Emirates.

* Correspondence: Huzaifa.umar@neu.edu.tr

Abstract

Cancer is one of the major causes of death in the modern world, and the incidence varies considerably based on race, ethnicity, and region. Novel cancer treatments, such as surgery and immunotherapy, are ineffective and expensive. In this situation, ion channels responsible for cell migration have appeared to be the most promising targets for cancer treatment. This research presents findings on the organic compounds present in *Albizia lebbbeck* ethanolic extracts (ALEE), as well as their impact on the anti-migratory, anti-proliferative and cytotoxic potentials on MDA-MB 231 and MCF-7 human breast cancer cell lines. In addition, artificial intelligence (AI) based models, multilayer perceptron (MLP), extreme gradient boosting (XGB), and extreme learning machine (ELM) were performed to predict *in vitro* cancer cell migration on both cell lines, based on our experimental data. The organic compounds composition of the ALEE was studied using gas chromatography-mass spectrometry (GC-MS) analysis. Cytotoxicity, anti-proliferations, and anti-migratory activity of the extract using Trypan Blue, MTT, and Wound Heal assay, respectively. Among the various concentrations (2.5~200 µg/mL) of the ALLE that were used in our study, 2.5~10 µg/mL revealed anti-migratory potential with increased concentrations, and they did not show any effect on the proliferation of the cells ($P < 0.05$; $n \geq 3$). Furthermore, the three data-driven models, Multilayer perceptron (MLP), Extreme gradient boosting (XGB), and Extreme learning machine (ELM), predict the potential migration ability of the extract on the treated cells based on our

experimental data. Overall, the concentrations of the plant extract that does not affect the proliferation of the type cells used demonstrated promising effects in reducing cell migration. XGB outperformed the MLP and ELM models and increased their performance efficiency by up to 3% and 1% for MCF and 1% and 2% for MDA-MB231, respectively, in the testing phase.

Keywords: Cell Migration; Cytotoxicity; Multilayer Perceptron; Extreme Gradient Boosting; Extreme Learning Machine.

Introduction

Cancer is a primary cause of death in the modern world, and the incidence varies considerably based on race, ethnicity, and region. Several studies revealed metastases responsible for 90% of cancer deaths [1,2,3,4]. Breast cancer (BCa) is among the leading causes of death in the United States and other parts of the world [5]. A recent report by the WHO (2018) shows that breast cancer accounts for about 0.627 million women's mortality. However, BCa patients' mortality rate increases from systematic metastases to distant organs [5].

Therefore, there is a need for effective chemo-preventive agents with fewer side effects to suppress tumour metastasis. BCa metastasis involves complex processes established by various pathways and factors. It starts with cell motility in the primary affected site to other distant tissue, blood, or lymph vessels [6,7]. Metastasis involves exiting cancerous cells from the primary site to invade various organs and tissues, facilitating cellular migration, invasion, and adhesion mediated by the Ca^{2+} signalling process [8]. Metastasis maintained primary tumour-differentiated characteristics such as cell-to-cell contact, signalling, and behaviour. Transforming Growth Factor (TGF) and Epidermal Growth Factor (EGF) signalling can trigger breast cancer tumour motility [9,10]. Consequently, understanding the mechanisms behind BCa metastasis and combatting it is paramount in defeating the war against BCa [11].

Therapies and drugs are available to cure cancer, but still, there is a need for effective therapies and targeted plant-based drugs with fewer side effects [12]. Various chemotherapeutic treatments with paclitaxel and anthracyclines might induce apoptosis, inhibit cell proliferation, and affect cellular activity [13]. Natural products such as plants, marine organisms, and microorganisms have revealed potential in cancer treatments [14]. There is a continuous search for anticancer drugs by scientists which will have low toxicity and side effects but high efficacy. One such approach is targeting apoptotic cells, usually targeted as a step during

therapy. Novel cancer treatments such as surgery and immunotherapy are not obligatory and quite expensive. In this situation, ion channels responsible for cell migration (metastasis) have appeared to be the most promising targets in the last two decades [15,14].

Albizia lebbbeck leaves are among those natural products rich in flavonoids and demonstrate antitumor activity against HepG2 hepatoma cells. Zinc oxide nanoparticles synthesized using plant stem bark revealed cytotoxic activity against highly and weakly metastatic human BCa cells [15,16]. Quercetin is one such flavonoid that is not toxic, and it demonstrates many biological actions, including anticancer, anti-inflammatory, and antioxidant activity [4,18,19]. Quercetin has also induced cell cycle arrest and apoptosis through signalling pathways and FOXO3a modification in breast cancer cells [20].

Modelling cancer cell migration *in vitro* has been a venerable challenge due to various complexities in cellular and molecular regulatory mechanisms in the system [21]. Conventionally, studies report the employment of rule-based greedy algorithms within agent-based modelling (ABM) in the modelling of cell migration [22,23]. Zhang et al. describe metastasis as the primary cause of death in many breast cancer patients. The cancer cell migration behaviour was analyzed computationally by applying high-content imaging and microfluidic single-cell migration. Random forest decision and artificial neural network (ANN) were employed in the study. The results indicated higher accuracy regarding the prediction of cell movement [24]. Ravdin et al. described the application of artificial intelligence for predicting clinical results of node-positive breast cancer patients. The status of the hormone receptor, tumour size, age of the patient, as well as relapse status, were used as the input variables. The results prove a satisfactory neural network in predicting cancer status in patients [25]. Furthermore, Jerez et al. employed various machine learning and statistical methods that can be used in the simulation of the recurrence of breast cancer patients. The results prove the reliability of the machine learning data algorithms over the classical statistical processes in the simulation of breast cancer outcomes [26]. Based on the studies mentioned above from the technical literature, it can be observed that the applications of artificial intelligence are of paramount importance, which shows reliable and satisfactory results over the classical statistical methods. Moreover, since the developments of artificial intelligence-based models, this is the first work conducted in the technical literature depicting the application of these kinds of novel artificial intelligence models for the prediction of lateral motility in human BCa cells in the technical literature.

In this present study, the organic compounds present in *Albizia lebbbeck* ethanolic extracts (ALEE), as well as their impact on the

on anti-migratory, antiproliferative and cytotoxic potential on MDA-MB 231 and MCF-7 human BCa cell lines. In addition, the artificial intelligence (AI) based models, multilayer perceptron (MLP), extreme gradient boosting (XGB), and extreme learning machine (ELM), were performed to predict *in vitro* cancer cells migration on MDA-MB 231 and MCF-7 cells, based on our experimental data.

Materials and Methods

2.1 Plant Material

Fresh stem barks of *A. lebbeck* were collected from northern Nigeria, a town called Tabuli, part of Gaya Local Government, Kano State, during their flowering stage and dried at room temperature. The collection of *A. lebbeck* stem bark follows all the applicable international standards guidelines and laws. The plant specimen was authenticated and deposited with voucher specimen number BUKHAN187 at the herbarium Plant Biology Department, Faculty of Science, Bayero University Kano.

2.2 Sample Preparation

Dried *Albizia lebbeck* stem barks were pulverized to clear powder and subjected to flask extraction using 99.9% methanol as extraction solvent. Powdered *A. lebbeck* stem bark (50g) was soaked in an Erlenmeyer flask containing methanol (500 mL) and placed under continual shaking for 48 h at room temperature [27]. Whatman filter paper No.1 was used to filter the extract and concentrate it under reduced pressure using a Rotary evaporator. The concentrated extract was dried completely at 40 °C in an oven and stored at 4 °C prior to the analysis.

2.3 Phytochemical Analysis of the Extract

The ALEE extracts were analyzed for their total flavonoid (TFC) and total phenolic content (TPC) using standard spectrophotometric methods [28,29]. To determine TFC determination, ALEE (1mg/mL) was mixed with NaNO₂ solution (5%), 10% AlCl₃, and 1 M NaOH, and absorbance was measured at 510 nm. Folin-Ciocalteu reagent was added to ALEE (10:1) for TPC, followed by incubation with Na₂CO₃ (7.5%) and measurement of absorbance at 760 nm. Results are presented as mg quercetin equivalent (QE)/g dry extract and gallic acid equivalents (µg GAEs/g dry extract).

We utilized gas chromatography-mass spectrometry (GC-MS) to analyze the organic composition of ALEE. To begin, we created a crude extract in ethanol (1mg/mL) and filtered it via a 0.22µm syringe filter. Then, we injected it into a Shimadzu GC-MS-QP2010 plus analyzer with helium as the carrier gas at a steady flow rate of 1mL/min. The initial oven temperature was set at 50 °C for 2min and gradually increased by 7 °C/min. We conducted an assessment of the mass spectra at a scanning interval of 0.5 seconds, with a complete scan range from 25m/z to 1000m/z, employing a Quadrupole mass detector. Ultimately, we identified the existing compounds by scrutinizing the spectrum via the WILLEY7 MS library.

2.3 Cell Models and Culture Conditions

MDA-MB 231 (strongly metastatic) and MCF-7 (weakly metastatic) BCa cell lines obtained as a gift from Imperial College London (UK) and stored at Biotechnology Research Centre (BCR) of Cyprus International University. The BCR ethical committee (BRCEC2011-01) approved the use of these cell lines in our study. We cultured the cells in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco by Life Technology USA), supplemented with 2mM L-glutamine, penicillin, and 10% fetal bovine serum (FBS), and maintained them in a sterile incubator at 37°C and 5% CO₂.

Cell lines used in our study were approved by BCR ethical committee (BRCEC2011-01). The cells lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Gibco by Life Technology USA) supplemented with 2mM L-glutamine, penicillin and fetal bovine serum (FBS) (10%), and penicillin. The cells were maintained in the sterile incubator at 37 °C and 5% CO₂.

2.4 Toxicity and Proliferation Assay

We conducted a trypan blue dye exclusion assay, following the guidelines provided by Fraser et al. [31], to measure the level of cytotoxicity in BCa cells. We administered various doses, ranging from 0 to 200 µg/mL, to the cells and observed them for 24, 48, and 72 hours. After this period, we replaced the medium with a diluted trypan blue solution, formulated by mixing 0.25 ml of the dye with 0.8 ml of medium. This assay accurately determined the extent of cytotoxicity present in the cells. Data are presented as averages of 3x30 measurements.

The proliferation of MDA-MB 231 (strongly metastatic) and MCF-7 (weakly metastatic) BCa cells treated with ALE extracts were assessed using MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) reagent Sigma-Aldrich) as described by Fraser et al. (1990) with some adjustments. BCa cells (3×10^4 cells/mL) cultured in tissue plates (12-well) were treated with 10, 5, 2.5, 0 $\mu\text{g/mL}$ of ALM and ALE extracts and incubated for 24, 48, and 72 hours. Treatments and culture medium (DMEM) were replaced every 24 hours. Microplate Reader (ELX 800™) was used to measure the absorbance of the treated cell and control at 570nm. All the experiment was performed at least thrice in triplicates ($n \geq 3$).

2.5 Wound Heal Assay

A wound heal assay was carried out to evaluate the anti-metastatic potential of ALM and ALE extracts against highly metastatic (MDA-MB 231) and weakly metastatic (MCF-7) cells using the method of Fraser *et al.* with some modifications. Cells were plated in 35 mm culture dishes, and parallel and intersecting lines were drawn on the culture dishes [31]. Briefly, $1 \times 10^6/\text{mL}$ and $5 \times 10^5/\text{mL}$ cells per dish of MCF-7 and MDA-MB 231, respectively, were plated on 35 mm culture dishes, and three scratch lines were made using pipette tips (200 μL) after the cell settled. The initial and subsequent wound made were captured using a camera (Leica, Germany) attached to an inverted microscope at x100 magnification, and image processing software (ImageJ) was used to analyse the recovery wound area (cell migration) by migrating cells using equation 1.

$$\text{Mo I} = 1 - \left(\frac{W_t}{W_0} \right) \quad (1)$$

Mo I, motility index; W_t , the width of the wound at 24 or 48 h; W_0 , initial wound width at 0 h.

2.6 Modelling Approach

The study of the science of data is very important in any data driven-model. The accuracy of the data was tested using XGB, ELM, and MLP algorithms with MATLAB (R2021a). In this work, various models were proposed for modelling the *in vitro* cancer metastasis prediction in MDA-MB 231 and MCF-7 cells, respectively. The data was collected from our experimental data set ($n \geq 80$) to reveal the accuracy of the algorithms. In this way, two parameters were used as input variables, i.e. the motility index on the cells and the concentration of the extract,

respectively, though other variables can also be used in the simulation of *in vitro* cancer metastasis prediction in both cell lines. In addition, models provide details on the effectiveness of the treatment, and choosing a single model that can perform best in most circumstances is difficult for the predictors, but applying various ensemble models can reveal the best models that will fit the data. Determination of cell migration potentials in breast cancer cells treated with ALE extract using the motility index on the cells and the concentration of the extract as the input parameters were the main objective of our proposed method. The proposed flowchart of the models is shown in Figure 1.

2.7 Extreme Gradient Boosting (XGB)

The XGB algorithm is a commonly used model that is highly efficient with high reproducibility in analyzing and modelling data using various inputs and outputs. The method was first introduced and improved by Friedman *et al.* [32], and it plays an essential role in the classification and regression of data. Its application in extreme learning techniques is well-known and the technique [33]. The technique uses a precise set up best complex decision tree algorithm to reveal good performance and speed faster than the normal gradient algorithm [34]. XGB is a machine learning ensemble technique that works similarly to Random Forest; and is recognized by its classification and regression trees (CART) set. Furthermore, it is not the same with the decision tree (DT), whereby every leave carries an actual score, which aids in enriching those interpretations which cannot be defined using the DT. Algorithms have been used in modelling and prediction of data, and it has shown promising results. Due to this ensemble technique's wide application and excellent features, we use it to model and predict the anti-migratory potential of the cells. Given that CART $[(x_i, y_i) \dots TK(x_i, y_i)]$ is the training data set of the treated cells motility index represented as x_i to predict outcomes y_i and determined using K classification, as shown in Equation (2) [35]:

$$\hat{y} = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (2)$$

218 where f_k represents independent tree structure with cells motility index scores and F denotes
 219 the space of all CART. Optimization of the objective is given by Eq. 3 [35]:

$$obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^t \Omega(f_i) \quad (3)$$

220 The loss function is denoted l , and it estimates the difference between target y_i and predicted
 221 $\hat{y}_i$. The regularization function that penalizes the model to avoid over-fitting is denoted as Ω ,
 222 and f_i represents the simultaneous training loss function. Furthermore, the prediction value
 223 for t at step $\hat{y}_i^t$ [35]:

224 Prediction $\hat{y}$ at the t step can be expressed as

$$\hat{y}_i^t = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{t-1} + f_t(x_i) \quad (4)$$

225 Substituting the predicted value in Eq. 4. Equation 3 can be expressed as [36]:

$$obj^t = \sum_{i=1}^n (y_i - (\hat{y}_i^{t-1} + f_t(x_i)))^2 + \sum_{i=1}^t \Omega(f_i) \quad (5)$$

226 It can also be expressed as

$$obj^t = \sum_{i=1}^n [2(\hat{y}_i^{t-1} - y_i)f_t(x_i) + f_t(x_i)^2 + \Omega(f_t) + constant] \quad (6)$$

227 Looking at Taylor's expansion due to loss of function, it can be expressed in Eq. (7) [36]:

$$obj^t = \sum_{i=1}^n [l(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 + \Omega(f_t) + constant] \quad (7)$$

228 where $g_i = \partial_{\hat{y}_i^{t-1}} l(y_i - \hat{y}_i^{t-1})$, and $h_i = \partial_{\hat{y}_i^{t-1}}^2 l(y_i - \hat{y}_i^{t-1})$. Which was the described by

229 $f_t(x) = w_{q(x)}$, and the normalized function is expressed as

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (8)$$

Where T represent the total number of trees and the objective function can rewritten as

$$\begin{aligned} obj^t &\approx \sum_{i=1}^n [g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\ &= \sum_{j=1}^T [(\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] + \gamma T \end{aligned} \quad (9)$$

where $I_j = \{i | q(x_i) = j\}$ refers to the j^{th} leaf data index. $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$, the objective function can be written as

$$obj^t = \sum_{j=1}^T [G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2] + \gamma T \quad (10)$$

Performance for $q(x)$, can be achieved by using the objective function and w_j as you can see in Eq. 11 and 12.

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (11)$$

$$obj^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (12)$$

In addition, Eq. (13) is for leaf node score during splitting, L and R are the left and right score, and the regularization of the additional leaf is denoted as γ .

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (13)$$

237

2.8 Extreme Learning Machine

The ELM model is a novel learning algorithm with a single hidden layer that works in a similar manner to a feed-forward neural network (FFNN) due to its approximation potential. And it was first introduced by Huang et al. [37]. Issues such as slower training speed and over-fitting present

with FFNN have been addressed analytically by ELM through inversion and matrix multiplication [38]. In addition, ELM hidden biases and input layer are chosen randomly, and the Moore-Penrose generalized inverse function determines the output layer. The ELM revealed precision due to its robustness when applied to hydrological Modelling [39].

The ELM was expressed by training dataset $\{(x_1, y_1), \dots, (x_t, y_t)\}$. Overall, the input are represented as $x_1, x_2, \dots, x_t$ and the output as $y_1, y_2, \dots, y_t$.

The training dataset N ($t = 1, 2, \dots, N$) where $x_t \in \mathbb{R}^d$ and $y_t \in \mathbb{R}$, with H hidden nodes, is given by [37] as in Eq. 14:

$$\sum_{i=1}^H B_i g_i(\alpha_i \cdot x_t + \beta_i) = z_t, \quad (14)$$

From equation (14) i represent index of the hidden layer node, β_i and α_i denotes the bias and weight of the random layers and d is the number of the inputs. Furthermore, the predicted weight of the output layer, model output and hidden layer neurons activation function, are $B \in \mathbb{R}^H$, $Z(z_t \in \mathbb{R})$ and $G(\alpha, \beta, x)$ respectively. The best activation function is found to be the sigMoId function [40] as follows:

$$G(x) = \frac{1}{1 + \exp(-x)}, \quad (15)$$

In addition, The output layer utilizes a linear activation function, which is shown in the following equation:

$$\sum_{t=1}^N \|z_t - y_t\| = 0, \quad (16)$$

The value of B is calculated using the system of linear equations as expressed in Eq. 17 and G in Eq. 18

$$Y = GB \quad (17)$$

$$G(\alpha, \beta, x) = \begin{bmatrix} g(x_1) \\ \vdots \\ g(x_N) \end{bmatrix} = \begin{bmatrix} g_1(\alpha_1 \cdot x_1 + \beta_1) & \cdots & g_L(w_H \cdot x_1 + \beta_H) \\ \vdots & \cdots & \vdots \\ g_1(\alpha_N \cdot x_N + \beta_1) & \cdots & g_L(w_H \cdot x_N + \beta_F) \end{bmatrix}_{N \times H} \quad (18)$$

261 B is calculated in Eq. 19, and Y in Eq. 20.

$$B = \begin{bmatrix} B_1^T \\ \vdots \\ B_H^T \end{bmatrix}_{H \times 1} \quad (19)$$

262

$$Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times 1} \quad (20)$$

263 G is for the hidden layer. $\hat{B}$ was calculated using “Moore-Penrose inverse function + by inverting
264 the hidden-layer matrix” (see, Eq. 21).

$$\hat{B} = G^+Y \quad (21)$$

265 Overall, estimated $\hat{y}$ which denotes the predicted MoI of the cells which can achieved using
266 Eq.22.

$$\hat{y} = \sum_{i=1}^H \hat{B}_i g_i(\alpha_i \cdot x_t + \beta_i) \quad (22)$$

267 **2.9 Multilayer Perceptron**

268 MLP, as one of the commonly applied Artificial neural networks (ANNs) composed of
269 information processing units together with an advanced tool used in simulation, motivated and
270 mimicked the biological neurons. In this way, ANN, just like the human central nervous system
271 (CNS), is capable of solving complex problems having a non-linear and linear behaviour
272 through combining features such as parallel processing, generalization, learning power and
273 decision making [41]. The general architecture of ANN consists of 3 layers with individual and
274 different tasks: The input layer, which acts in distributing the data in the network, the hidden
275 layers, which process the information and the outputs, which, in addition to processing each
276 input vector, show its output. The neurons are regarded as the smallest unit that processes the
277 networks. The basic characteristics of MLP include the use of interactive connections between
278 the neurons without advanced mathematical design to complete the information processing.

Furthermore, MLP is composed of input, one or more hidden and output layers in its architecture, similar to the ANN (Fig. 2) [40].

2.10 Performance objectives

To evaluate the performance efficiency of the artificial intelligence based models used in the current study; two different metrics, where; Nash–Sutcliffe coefficient (NS) was used for understanding the fitness between the experimental and predicted values, while, Root mean square error (RMSE) was used in determining the errors depicted by each model.

Hence, Root mean square error (RMSE) was expressed as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N ((Y)_{obs,j} - (Y)_{com,j})^2} \quad (23)$$

Nash–Sutcliffe coefficient (NS), expressed as:

$$NS = 1 - \left[\frac{\sum_{i=1}^N (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^N (Q_{obs,i} - \bar{Q}_{obs,i})^2} \right] \infty \leq NS \leq 1 \quad (2)$$

3. Result and Discussion

3.1 Experimental Results

The study found that the ALEE contained TFC and TPC at levels of 2022.80 ± 17.83 QE $\mu\text{g/g}$ and 6556.49 ± 22.52 GAE $\mu\text{g/g}$, respectively. Studies have shown that TPC is highly efficient in scavenging different oxidizing molecules, including free radicals produced during lipid peroxidation [42]. Moreover, research has revealed that flavonoids, which are present in various structures of phenolic compounds, possess medicinal properties. These compounds can be found in sources such as flowers, leaves, stem bark, roots, fruits and tea [43, 44].

The compounds found in ALEE are listed in Table 1, their corresponding chromatogram peaks are shown in Figure 3. We identified several major compounds in our extract that have biological potential; some of them are; Ethanol (88.55%), Silicic acid, diethyl bis(trimethylsilyl) ester (3.18%), 1-cyano-5-benzoyloxy- α -D-ribofuranose (1.67%), 1-(2-trimethylsiloxy-1,1-dideuteriovinyl)-4-trimethyl siloxy-benzene (1.21%), Disiloxane, 1,3-

diethoxy-1,1,3,3-tetramethyl- (1.43%), 1,2-Dihydro-1,4-diphenylphthalazine (0.75 %), 3-(4'-Methoxyphenyl)-1-acetyl-2-phenylindolizine (0.52%), 4H-3-(p-methylamino)1-benzothiopyran-4-one 1-oxide (0.40%). The high percentage of ethanol might be from the extraction solvent, and it indicates that ethanol is not a suitable solvent for *A. lebbeck* extraction. Furthermore, the anti-proliferative and anti-migratory potential of the extract might be a result of the phytochemicals present in the extract, and ALEE could be a good cause that will prevent the metastasis of breast cancer.

The effect of various concentrations (2.5~200 µg/mL) of ALEE on human BCa cells for 24h and 48 hours, and cytotoxicity and effect on proliferation were determined using trypan blue assay and MTT, respectively (Figures 4 and 5). Various ALEE concentrations used in the study are 2.5, 5 and 10 µg/mL, and they demonstrated no effect on the viability of both cells, but treatment with a concentration between 25~200 µg/mL revealed significant changes ($P < 0.05$). Treatment of MDA-MB 231 with 2.5, 5 and 10 µg/mL ALEE did not significantly alter cell viability when compared with untreated cells (control). Similarly, treatment of MCF-7 with 2.5, 5 and 10 µg/mL ALEE did not show significant changes compared with the control ($P > 0.05$). Studies revealed *in vitro* anti-proliferative potential of Silicic acid, diethyl bis(trimethylsilyl) ester separated from *Lorabthus parasiticus* on breast cancer in a dose-dependent manner, which is in agreement with our findings [49]. (S)-(E)-(-)-4-Acetoxy-1-phenyl-2-dodecen-1-one (Quercetin) isolated from green tea revealed anti-proliferative effect against in PC-3 and LNCaP human prostate cancer cells [50]. Furthermore, studies revealed that quercetin isolated from plants inhibits proliferation, signal transduction and metastasis in cancer cell lines [51].

According to the study, the concentration of ALEE did not have a notable impact on the viability and growth of MDA-MB 231 and MCF-7 human BCa cells when compared to the control group. Nonetheless, the anti-migratory capacity of the cells was examined through the wound healing assay, and it was discovered that the lateral motility index (MOI) of MDA-MB 231 decreased with an increase in ALEE concentration and incubation duration. Figure 6 indicates that 10 µg/mL of ALEE had the most optimal motility index among the other concentrations. The MOI is more in MDA-MB 231 because the cells are powerfully metastatic and aggressive. In addition, all ALEE concentrations revealed significant differences relative to the control ($P < 0.05$) (Figure 6). Similarly, the MCF-7 MOI was reduced with increased ALEE concentration and incubation period, as shown in Figure 6 d-e, and 10 µg/mL revealed the lowest and best MOI when compared with the remaining ALEE concentrations. MCF-7 is a less aggressive and weakly metastatic cell, and that could be the reason for the lower MoI in

comparison with MDA-MB 231 cells. The ALEE concentrations (2.5~10 µg/mL) revealed a decrease in the MOI of MCF-7 cells with increased concentrations and incubation time, and 10 µg/mL revealed more effect on lateral motility followed by 5 µg/mL (Figure d-e; $n \geq 3$). Nanoparticles synthesized using quercetin reached plant (*Ficus ingens*) revealed an effect on lateral motility of MDA-MB 231 [13]. Medicinal plants containing quercetin as an active ingredient showed anti-metastatic activity on strongly and weakly metastatic MatLYLu and AT-2 rat prostate cancer cell models, respectively [52].

3.2 Anti-migratory Potential Prediction Models

The AI-based models (MLP, XGB, and ELM) were analysed to predict *in vitro* cancer migration prediction in cells treated with ALEE based on our experimental data. Before the model calibration, statistical data analysis was conducted, as shown in Table 2. Generally, statistical analysis is done to understand the dataset. Furthermore, The AI-based models (MLP, XGB, and ELM) were analysed to predict *in vitro* cancer migration prediction in the MDA-MB 231 and MCF-7 human BCa, treated with ALEE based on our experimental data. The performance evaluation is checked by applying various criteria based on comparing the simulated and the experimental values. The distribution between the various parameters and the dataset used in the study was expressed in the form of visualised pie chart in Figure 7, and the data set is well distributed. Furthermore, the correlation matrix shows the correlation that exists between different parameters in a linear form. It can be seen from Figure 8 that there is a high correlation between all the parameters, whereby the highest correlation in this study is between MD-AMB231 and MCF-7 having R-value =0.98, and the lowest correlation exists between concentration and MCF-7 with R=0.75. Similarly, the correlation matrix proves to have shown a robust correlation between all the variables and is in conformity with the correlation revealed by Adun et al. [53].

The modelling performance of MLP, XGB, and ELM models, treated MDA-MB 231 and MCF-7, were compared to each other using RMSE and NSE, as shown in Table 2. Based on

the predictive comparison of the models in Table 3, it can be shown clearly that all three data-driven models (MLP, XGB and ELM) have the ability to simulate the *in vitro* cancer migration potential prediction in the human BCa cells. XGB depicted the superiority over the other two non-linear models in the testing and training stages for modelling the performance of the cells. In regards to their error values, XGB shows the lowest RMSE values, XGB-MCF-7= 0.0039 and XGB-MDAMB231=0.0025 in the testing phase, and the NSE as a goodness of fits which shows that XGB equally outperformed all the other AI-based models MLP and ELM and increase their performance efficiency up to 3% and 1% for MCF and 1% and 2% for MDA-MB231 respectively in the testing phase. The relative predictive accuracy regarding the relative error can also be demonstrated using a bar chart (Figure 9), which demonstrates the performance of *in vitro* cancer metastasis prediction in human BCa cells in a surface radar chart showing the scale of NSE in the training and testing phases. It has been reported that the scale of radar generally ranges between 0 and 1. The radar chart performance demonstrated that the performances in terms of NSE of treated BCa cells migration in highly and weakly metastatic human BCa cell lines follow the following order XGB>ELP>ML for MCF-7, and XGB>MLP>ELP for MDA-MB 231, respectively (Figure10). Similarly, the methanolic extract of *A. lebbeck* demonstrated good performance using other AI-based models [54].

Conclusion

Our study has uncovered promising organic compounds in ALEE that possess medicinal properties, potentially aiding in the prevention of metastasis in human breast cancer. Interestingly, we observed that varied concentrations of the plant extract were non-toxic and had no impact on cell proliferation, but displayed significant anti-migratory potential in both MDA-MB 231 and MCF-7 cells, with increasing concentration. Furthermore, we found that AI models, including MLP, XGB, and ELM, were effective in predicting the anti-migratory potential of ALEE. XGB demonstrated the highest performance efficiency, outperforming MLP and ELM models by 3% and 1% for MCF and 1% and 2% for MDA-MB231, respectively, during the testing phase. However, further studies are required to validate the anti-migratory potential of this plant, and additional computational models should be employed to improve performance.

395 **Data Availability**

396 All data is included in the manuscript

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

References

- [1] Bogenrieder, T., & Herlyn, M. (2003). Axis of evil: Molecular mechanisms of cancer metastasis. *Oncogene*. <https://doi.org/10.1038/sj.onc.1206757>.
- [2] Dillekås, H., Rogers, M. S., & Straume, O. (2019). Are 90% of deaths from cancer caused by metastases? *Cancer Medicine*. <https://doi.org/10.1002/cam4.2474>.
- [3] Geiger, T. R., & Peeper, D. S. (2009). Metastasis mechanisms. *Biochimica et Biophysica Acta - Reviews on Cancer*. <https://doi.org/10.1016/j.bbcan.2009.07.006>.
- [4] Gupta, G. P., & Massagué, J. (2006). Cancer Metastasis: Building a Framework. *Cell*. <https://doi.org/10.1016/j.cell.2006.11.001>.
- [5] Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*. <https://doi.org/10.3322/caac.21551>.
- [6] Sahai, E. (2007). Illuminating the metastatic process. *Nature Reviews Cancer*. <https://doi.org/10.1038/nrc2229>.
- [7] Weigelt, B., Peterse, J. L., & Van't Veer, L. J. (2005). Breast cancer metastasis: Markers and models. *Nature Reviews Cancer*. <https://doi.org/10.1038/nrc1670>.
- [8] Penna, A., Juvin, V., Chemin, J., Compan, V., Monet, M., & Rassendren, F. A. (2006). PI3-kinase promotes TRPV2 activity independently of channel translocation to the plasma membrane. *Cell Calcium*. <https://doi.org/10.1016/j.ceca.2006.01.009>.
- [9] Jakowlew, S. B. (2006). Transforming growth factor- β in cancer and metastasis. *Cancer and Metastasis Reviews*. <https://doi.org/10.1007/s10555-006-9006-2>.
- [10] Lesko, E., & Majka, M. (2008). The biological role of HGF-MET axis in tumor growth and development of metastasis. *Frontiers in Bioscience*. <https://doi.org/10.2741/2760>.

- 437 [11] Giampieri, S., Manning, C., Hooper, S., Jones, L., Hill, C. S., & Sahai, E. (2009).
438 Localized and reversible TGF β signalling switches breast cancer cells from cohesive to
439 single cell motility. *Nature Cell Biology*. <https://doi.org/10.1038/ncb1973>.
- 440 [12] Singh, J., Hussain, Y., Luqman, S., & Meena, A. (2019). Targeting Ca²⁺ signalling
441 through phytochemicals to combat cancer. *Pharmacological Research*.
442 <https://doi.org/10.1016/j.phrs.2019.104282>.
- 443 [13] Kavaz, D., Umar, H., & Shehu, S. (2018). Synthesis, characterization, antimicrobial and
444 antimetastatic activity of silver nanoparticles synthesized from Ficus ingens leaf. *Artificial*
445 *Cells, Nanomedicine and Biotechnology*.
446 <https://doi.org/10.1080/21691401.2018.1536060>
- 447 [14] Leanza, L., Managò, A., Zoratti, M., Gulbins, E., & Szabo, I. (2016). Pharmacological
448 targeting of ion channels for cancer therapy: In vivo evidences. *Biochimica et Biophysica*
449 *Acta - Molecular Cell Research*.
- 450 [15] Djamgoz, M. B. A., Coombes, R. C., & Schwab, A. (2014). Ion transport and cancer: From
451 initiation to metastasis. *Philosophical Transactions of the Royal Society B: Biological*
452 *Sciences*. <https://doi.org/10.1098/rstb.2013.0092>.
- 453 [16] Lam, S. K., & Ng, T. B. (2011). First report of an anti-tumor, anti-fungal, anti-yeast and
454 anti-bacterial hemolysin from Albizia lebbeck seeds. *Phytomedicine*.
455 <https://doi.org/10.1016/j.phymed.2010.08.009>.
- 456 [17] Umar, H., Kavaz, D., & Rizaner, N. (2019). Biosynthesis of zinc oxide nanoparticles using
457 albizia lebbeck stem bark, and evaluation of its antimicrobial, antioxidant, and cytotoxic
458 activities on human breast cancer cell lines. *International Journal of Nanomedicine*, 14,
459 87–100. <https://doi.org/10.2147/IJN.S186888>.

- [18] da Silva, A. B., Cerqueira Coelho, P. L., das Neves Oliveira, M., Oliveira, J. L., Oliveira Amparo, J. A., da Silva, K. C., ... Costa, S. L. (2020). The flavonoid rutin and its aglycone quercetin modulate the microglia inflammatory profile improving antiglioma activity. *Brain, Behavior, and Immunity*. <https://doi.org/10.1016/j.bbi.2019.05.003>
- [19] Morel I, Lescoat G, Cogrel P, Sergent O, Padeloup N, Brissot P, ... Cillard J. (1993). Antioxidant and iron-chelating activities of the flavonoids catechin, quercetin and diosmetin on iron-loaded rat hepatocyte cultures. *Biochemical Pharmacology*.
- [20] Nguyen, L. T., Lee, Y. H., Sharma, A. R., Park, J. B., Jagga, S., Sharma, G., ... Nam, J. S. (2017). Quercetin induces apoptosis and cell cycle arrest in triple-negative breast cancer cells through modulation of Foxo3a activity. *Korean Journal of Physiology and Pharmacology*. <https://doi.org/10.4196/kjpp.2017.21.2.205>.
- [21] Kusumoto, D., & Yuasa, S. (2019). The application of convolutional neural network to stem cell biology. *Inflammation and Regeneration*. <https://doi.org/10.1186/s41232-019-0103-3>
- [22] Setty, Y., Cohen, I. R., Dor, Y., & Harel, D. (2008). Four-dimensional realistic modelling of pancreatic organogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0808725105>.
- [23] Wang, Z., Ramsey, B. J., Wang, D., Wong, K., Li, H., Wang, E., & Bao, Z. (2016). An observation-driven agent-based modelling and analysis framework for *C. elegans* embryogenesis. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0166551>
- [24] Zhang, Z., Chen, L., Humphries, B., Brien, R., Wicha, M. S., Luker, K. E., ... Yoon, E. (2018). Integrative Biology Morphology-based prediction of cancer cell migration using an artificial neural network and a random decision forest †. *Integrative Biology*, 10, 758–767. <https://doi.org/10.1039/c8ib00106e>.

- [25] Ravdin, P. M., Clark, G. M., Hilsenbeck, S. G., Owens, M. A., Vendely, P., & Mcguire, W. L. (1992). *A demonstration that breast cancer recurrence can be predicted by neural network analysis*. 47–53.
- [26] Jerez, M., Molina, I., Garcı, P. J., Alba, E., Ribelles, N., Franco, L., & Martı, M. (2010). *Artificial Intelligence in Medicine Missing data imputation using statistical and machine learning methods in a real breast cancer problem*. 50, 105–115. <https://doi.org/10.1016/j.artmed.2010.05.002>.
- [27] Umar, H., Kavaz, D., Abubakar, A. L., Aliyu, M. R., & Rizaner, N. (2022). Synthesis of zinc oxide nanoparticles using *Ficus thonningii* aqueous extract and evaluation of its anti-oxidant and anti-microbial activities. *BULGARIAN CHEMICAL COMMUNICATIONS*, 277.
- [28] Singleton, V. L., & Rossi, J. A. (1965). Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *American journal of Enology and Viticulture*, 16(3), 144-158.
- [29] Meda, A., Lamien, C. E., Romito, M., Millogo, J., & Nacoulma, O. G. (2005). Determination of the total phenolic, flavonoid, and proline contents in Burkina Fasan honey, as well as their radical scavenging activity. *Food Chemistry*, 91(3), 571-577.
- [30] Kumar, J., Dhar, P., Tayde, A. B., Gupta, D., Chaurasia, O. P., Upreti, D. K., Toppo, K., Arora, R., Suseela, M. R., Ravi B. and Srivastava, R.B. (2015). Chemical Composition and Biological Activities of Trans-Himalayan Alga *Spirogyra porticalis* (Muell.) Cleve. *Plos One*. 10(2).
- [31] Fraser SP, Salvador V, Manning EA, Mizal J, Altun S, Raza M, MBA D. (2003). Contribution of functional voltage-gated Na⁺ channel expression to cell behaviors involved in the metastatic cascade. *J Cell Physiol.*, 195 (3), 479–87.

- 508 [32] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine.
509 *Annals of Statistics*.
- 510 [33] Pradhan, B., & Sameen, M. I. (2020). Predicting Injury Severity of Road Traffic Accidents
511 Using a Hybrid Extreme Gradient Boosting and Deep Neural Network Approach. In *Laser*
512 *Scanning Systems in Highway and Safety Assessment* (pp. 119–127). Springer.
- 513 [34] Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings*
514 *of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data*
515 *Mining*, 785–794. ACM.
- 516 [35] Chen, Z.-Y., Zhang, T.-H., Zhang, R., Zhu, Z.-M., Yang, J., Chen, P.-Y., ... Guo, Y.
517 (2019). Extreme gradient boosting model to estimate PM2.5 concentrations with missing-
518 filled satellite data in China. *Atmospheric Environment*, 202, 180–189.
519 <https://doi.org/10.1016/J.ATMOSENV.2019.01.027>.
- 520 [36] Chen, T., He, T., Benesty, M., Khotilovich, V., & Tang, Y. (2015). Xgboost: extreme
521 gradient boosting. *R Package Version 0.4-2*, 1–4.
- 522 [37] Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and
523 applications. *Neurocomputing*, 70(1–3), 489–501.
524 <https://doi.org/10.1016/j.neucom.2005.12.126>.
- 525 [38] Huang, G., Huang, G. Bin, Song, S., & You, K. (2015). Trends in extreme learning
526 machines: A review. *Neural Networks*, 61, 32–48.
527 <https://doi.org/10.1016/j.neunet.2014.10.001>.
- 528 [39] Yaseen, Zaher M, Allawi, M. F., Yousif, A. A., Jaafar, O., Hamzah, F. M., & El-Shafie,
529 A. (2016). Non-tuned machine learning approach for hydrological time series forecasting.
530 *Neural Computing and Applications*, 1–13. <https://doi.org/10.1007/s00521-016-2763-0>

- [40] Fijani, E., Barzegar, R., Deo, R., Tziritis, E., & Konstantinos, S. (2019). Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. *Science of the Total Environment*, 648, 839–853. <https://doi.org/10.1016/j.scitotenv.2018.08.221>.
- [41] Ahmad, M.H., Usman, A.G. & Abba, S.I. (2021). Comparative performance of extreme learning machine and Hammerstein–Weiner models for modelling the intestinal hypermotility and secretory inhibitory effects of methanolic leaf extract of *Combretum hypopilinum* Diels (Combretaceae). In *Silico Pharmacol.* **9**, 31. <https://doi.org/10.1007/s40203-021-00090-1>.
- [42] Aryal S, Baniya MK, Danekhu K, Kunwar P, Gurung R, Koirala N. (2019). Total Phenolic Content, Flavonoid Content and Antioxidant Potential of Wild Vegetables from Western Nepal. *Plants (Basel)*. 11;8(4):96. doi: 10.3390/plants8040096.
- [43] Ayala A, Muñoz MF, Argüelles S. (2014). Lipid peroxidation: production, metabolism, and signaling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. *Oxid Med Cell Longev.* ;2014:360438. doi: 10.1155/2014/360438.
- [44] Tungmunthum D, Thongboonyou A, Pholboon A, Yangsabai A. (2018). Flavonoids and Other Phenolic Compounds from Medicinal Plants for Pharmaceutical and Medical Aspects: An Overview. *Medicines (Basel)*. 25;5(3):93. doi: 10.3390/medicines5030093.
- [45] Kampf, G., Hollingsworth, A. (2008). Comprehensive bactericidal activity of an ethanol-based hand gel in 15 seconds. *Ann Clin Microbiol Antimicrob* **7**, 2. doi.org/10.1186/1476-0711-7-2
- [46] Musa MA, Badisa VL, Latinwo LM, Cooperwood J, Sinclair A, Abdullah A. (2011). Cytotoxic activity of new acetoxycoumarin derivatives in cancer cell lines. *Anticancer*

Res. 31(6):2017-22.

[47] Shin, S.Y., Park, J., Jung, Y. *et al.* (2020). Anticancer activities of cyclohexenone derivatives. *Appl Biol Chem* **63**, 82 (2020). <https://doi.org/10.1186/s13765-020-00567-1>.

[48] Dessie Ezez, Natinael Mekonnen & Molla Tefera (2023) Phytochemical analysis of *Withania somnifera* leaf extracts by GC-MS and evaluating antioxidants and antibacterial activities, *International Journal of Food Properties*, 26:1, 581-590, doi:[10.1080/10942912.2023.2173229](https://doi.org/10.1080/10942912.2023.2173229).

[49] Devi, Y. P., Uma, A., Narasu, M. L., & Kalyani, C. (2014). Anticancer Activity of Gallic Acid on Cancer Cell Lines, HCT15 and MDA MB 231. *International Journal of Research in Applied, Natural and Social Sciences*.

[50] Wang, P., Heber, D., & Henning, S. M. (2012). Quercetin increased the antiproliferative activity of green tea polyphenol (-)-epigallocatechin gallate in prostate cancer cells. *Nutrition and Cancer*. <https://doi.org/10.1080/01635581.2012.661514>.

[51] Vidya Priyadarsini, R., Senthil Murugan, R., Maitreyi, S., Ramalingam, K., Karunagaran, D., & Nagini, S. (2010). The flavonoid quercetin induces cell cycle arrest and mitochondria-mediated apoptosis in human cervical cancer (HeLa) cells through p53 induction and NF- κ B inhibition. *European Journal of Pharmacology*.

[52] Gumushan-Aktas, H., & Altun, S. (2016). Effects of Hedera helix L. extracts on rat prostate cancer cell proliferation and motility. *Oncology Letters*. <https://doi.org/10.3892/ol.2016.4941>

[53] Adun, H., Kavaz, D., Dagbasi, M., Umar, H., & Wole-Osho, I. (2021). An experimental investigation of thermal conductivity and dynamic viscosity of Al₂O₃-ZnO-Fe₃O₄ ternary hybrid nanofluid and development of machine learning model. *Powder*

Technology, 394, 1121-1140.

- [54] Umar, H.; Rizaner, N.; Usman, A.G.; Aliyu, M.R.; Adun, H.; Ghali, U.M.; Uzun Ozsahin, D.; Abba, S.I. Prediction of Cell Migration in MDA-MB 231 and MCF-7 Human Breast Cancer Cells Treated with *Albizia Lebbeck* Methanolic Extract Using Multilinear Regression and Artificial Intelligence-Based Models. *Pharmaceuticals* **2023**, *16*, 858. <https://doi.org/10.3390/ph16060858>.

598 **Table 1:** Bioactive compounds identified from ALEE by GC-MS analysis and their biological
599 activities.

S/n	RT	Peak Area	Area %	Compound detected	Biological Activity
1	2.00	25065811 ⁵⁶	88.55	Ethanol (CAS)	Anti-microbial [45]
2	3.08	2369689	0.08	(S)-(E)-(-)-4-Acetoxy-1-phenyl-2-dodecen-1-one	Anti-cancer, DNA-binding and anti-microbial [46]
3	3.77	13527363	0.48	Cyclotrisiloxane, hexamethyl- (CAS)	Anti-cancer, anti-oxidant and anti-microbial [47]
4	3.98	21342385	0.75	1,2-Dihydro-1,4-diphenylphthalazine	Anticancer [46].
5	6.46	90085996	3.18	Silicic acid, diethyl bis(trimethylsilyl) ester	-
6	6.74	34182822	1.21	1-(2-trimethylsiloxy-1,1-dideuteriovinyl)-4-trimethylsiloxy-benzene	Anti-microbial [45]
7	7.60	47188509	1.67	1-cyano-5-benzoyloxy- α -D-ribofuranose	Anticancer [46]
8	11.33	40503790	1.43	Disiloxane, 1,3-diethoxy-1,1,3,3-tetramethyl- (CAS)	Anti-inflammatory, anti-bacterial [48]
9	13.47	3392203	0.12	(S)-(E)-(-)-4-Acetoxy-1-phenyl-2-dodecen-1-one	Anti-cancer, DNA-binding and anti-microbial [46]
10	13.76	15872886	0.56	Silicic acid, diethyl bis(trimethylsilyl) ester	-
11	15.69	5409162	0.19	Silicic acid, diethyl bis(trimethylsilyl) ester	-
12	16.67	8613439	0.30	2-(5-acetyl-2-thienyl)-1,4-naphthoquinone	Anti-microbial, anti-tumor and antiproliferative [46]
13	16.84	11375011	0.40	4H-3-(p-methylanilino)1-benzothiopyran-4-one oxide	1- Anti-cancer [46].
14	17.29	3343231	0.12	Heptasiloxane, 1,1,3,3,5,5,7,7,9,9,11,11,13,13-tetradecamethyl-	Anti-oxidant and anti-inflammatory [48]
15	18.70	14832408	0.52	3-(4'-Methoxyphenyl)-1-acetyl-2-phenylindolizine	-
16	23.31	3451675	0.12	Cyclooctasiloxane, hexadecamethyl-	Anti-cancer, anti-oxidant and anti-microbial [47]
17	25.33	6562803	0.23	4-Ethylbenzoic acid, 2-butyl ester	Anti-cancer and Anti-microbial [47]
18	29.18	2201203	0.08	Cyclononasiloxane, octadecamethyl-	Anti-cancer, anti-oxidant and anti-microbial [48]
18	29.18	2201203	0.08	Cyclononasiloxane, octadecamethyl-	Anti-cancer, anti-oxidant and anti-microbial [46]

600
601
602
603
604

605

606

607

608

609

610 **Table 2:** Descriptive Statistic

<i>Variables</i>	Mean	SE	Median	SD	Kurtosis	Skewness	Min	Max
<i>Conc.</i>	4.375	0.48138	3.75	3.728753	-1.14951	0.445876	0	10
<i>MCF-7</i>	0.405333	0.017451	0.47	0.135177	-0.5607	-1.10911	0.12	0.54
<i>MDA-MB</i> <i>231</i>	0.402167	0.010146	0.435	0.078591	-0.5463	-1.04085	0.24	0.49

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

Table 3. Result of the ANN, ANFIS and MLR models.

	Calibration		Verification	
	NSE	RMSE	NSE	RMSE
MLP-MCF-7	0.9964	0.0263	0.9692	0.0302
XGB-MCF-7	0.9999	0.0033	0.9975	0.0039
ELM-MCF-7	0.9998	0.0052	0.9974	0.0064
MLP-MDAMB231	0.9955	0.0181	0.9871	0.0356
XGB-MDAMB231	0.9985	0.0018	0.9974	0.0025
ELM-MDAMB231	0.9924	0.0232	0.9728	0.0432

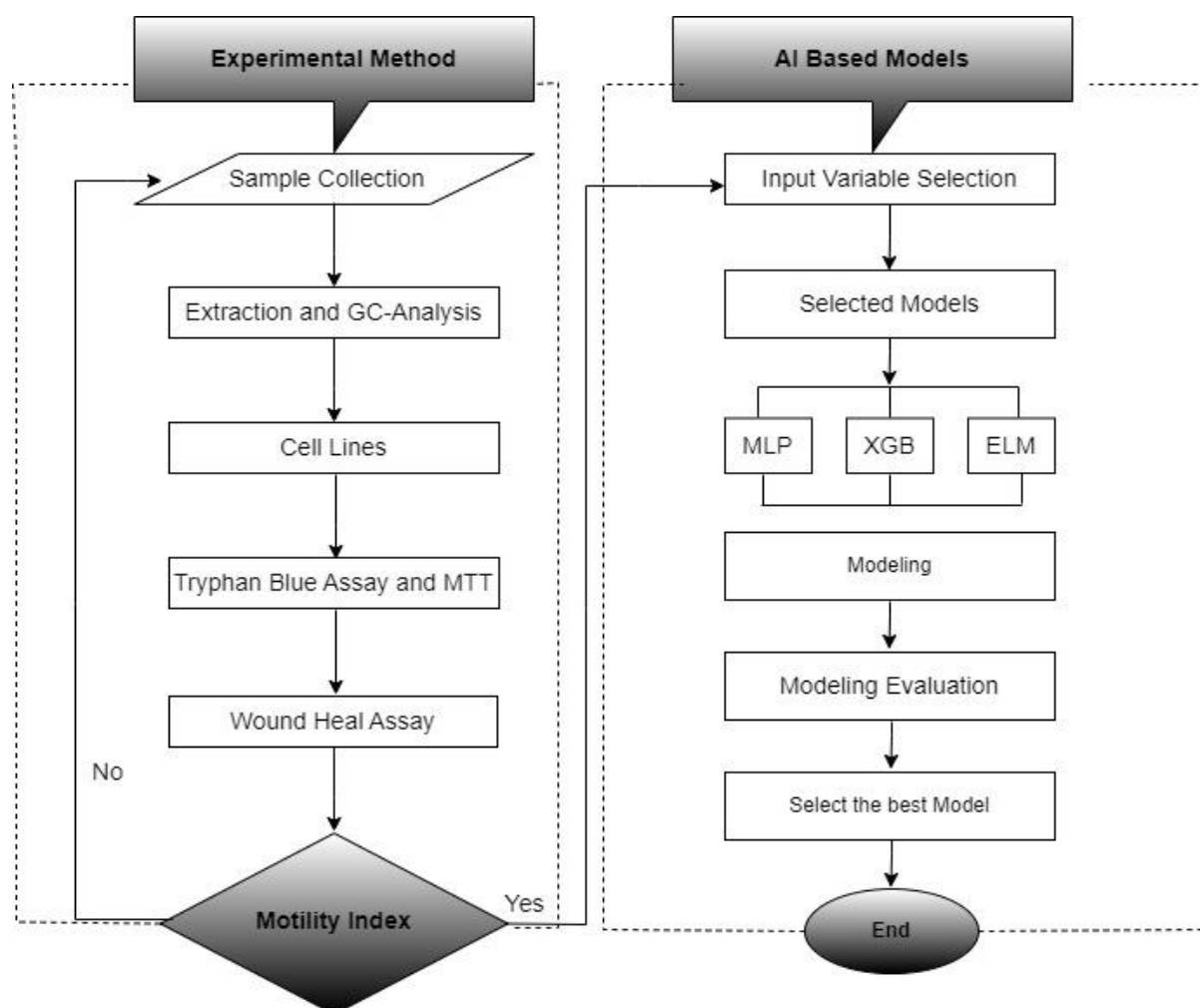


Figure 1. Proposed flowchart of experimental data-driven methods

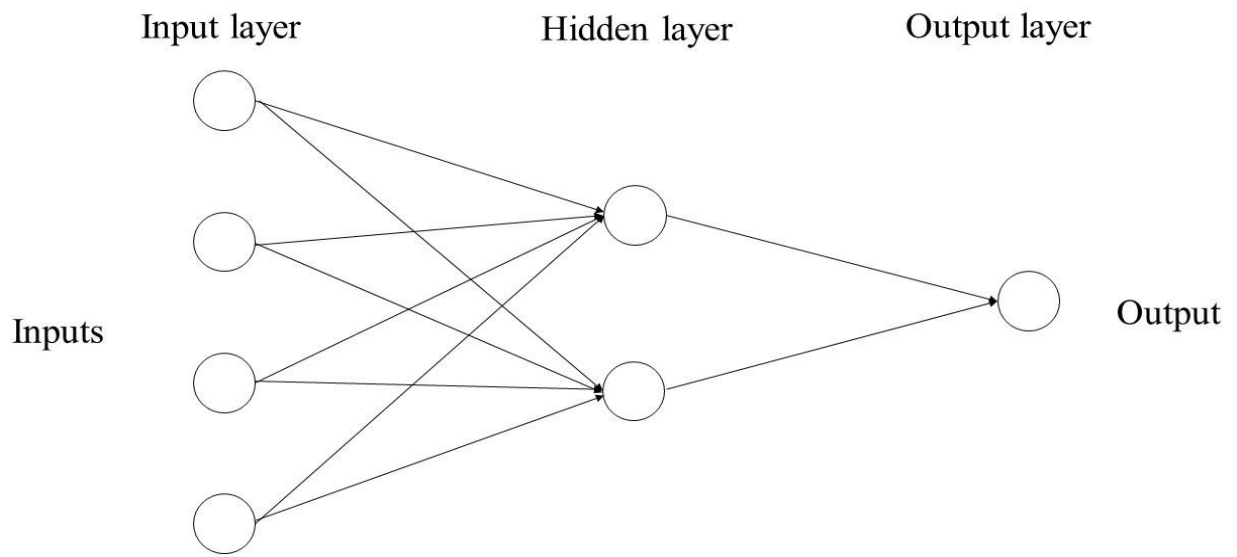


Figure 2. Schematic diagram of MLP network structure.

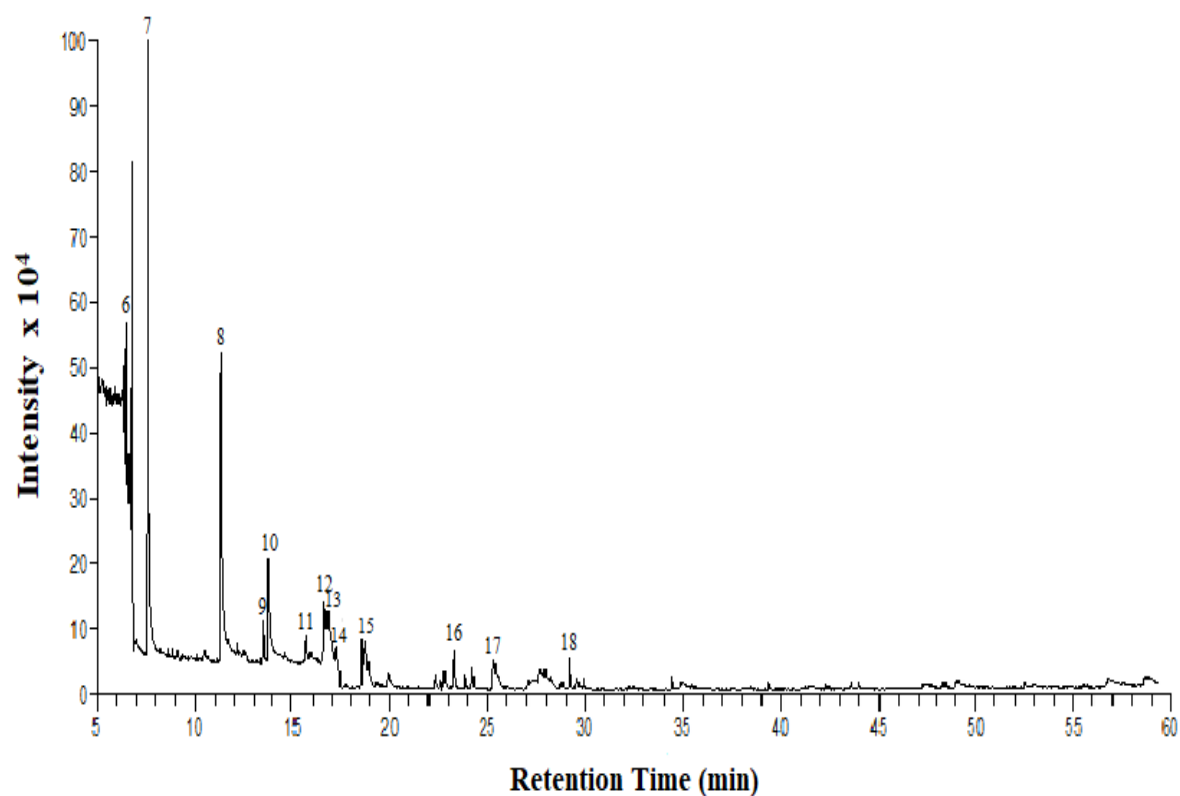


Figure 3: GC-MS chromatogram of ALEE bioactive compounds. The numbered peaks correspond to the numbers and molecules in Table 1.

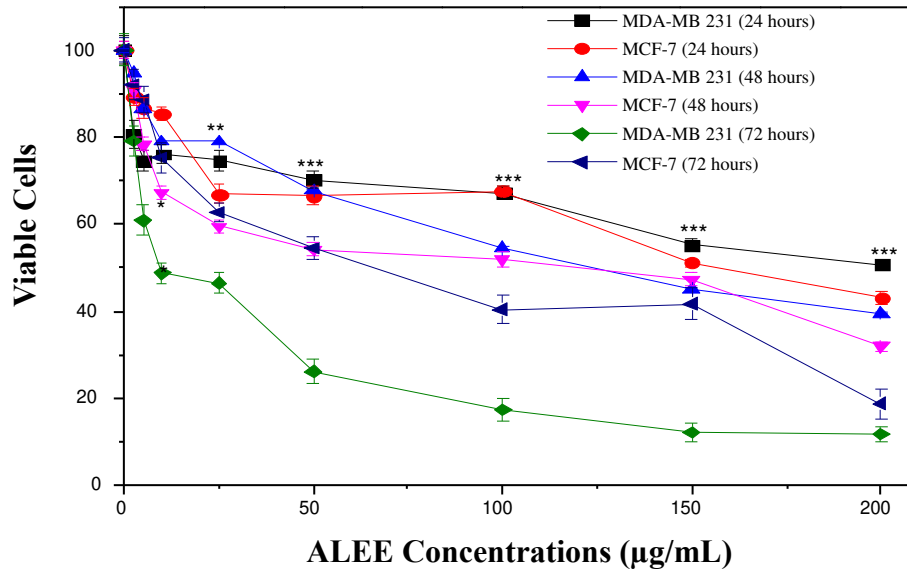
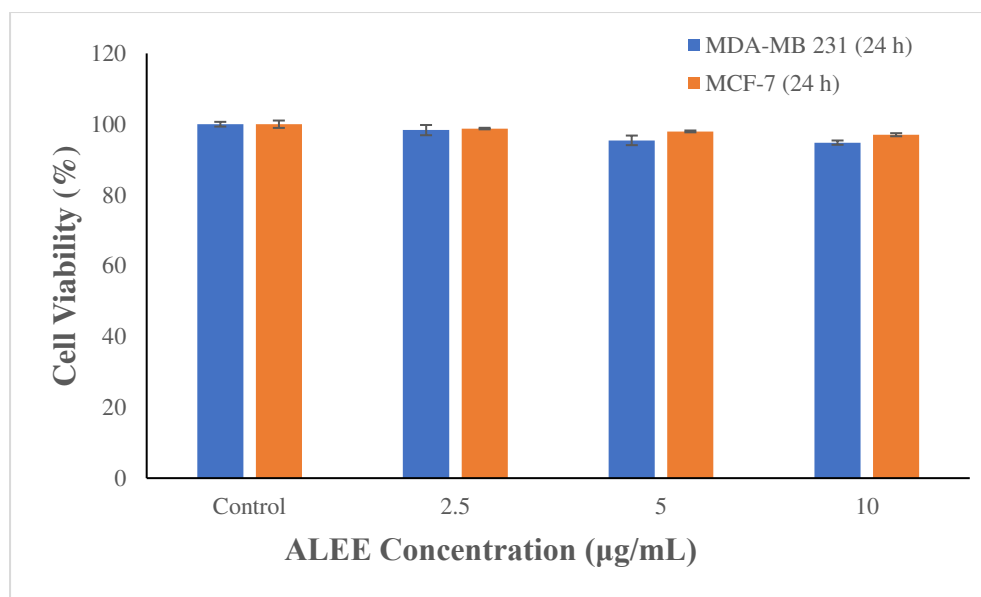


Figure 4. Cytotoxic Effect of various concentration of ALEE MDA-MB 231 and MCF-7 breast cancer cell lines carried out using tryphan blue exclusion assay. The observations were made after 24, 48 and 72 hours' incubation period. The findings are presented as mean $\pm$ SEM of at least replicates experiments ($n \geq 3$) and analysed using one-way ANOVA followed Newman-Keuls post hoc analysis. (*) $P < 0.05$; (**) $P < 0.01$ and (***) $P < 0.0001$.

(a)



(b)

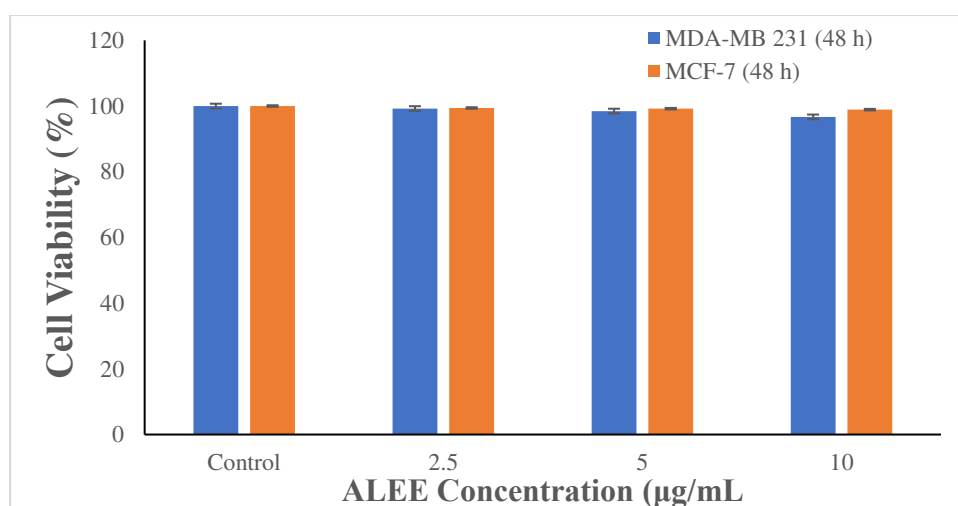


Figure 5. Effect of ALEE on proliferation of MDA-MB 231 and MCF. MDA-MB 231 and MCF-7 were treated with various concentration of ALEE for (a) 24 h and (b) 48 h. Data are presented as mean $\pm$ SEM of at least replicates experiments ($n \geq 3$).

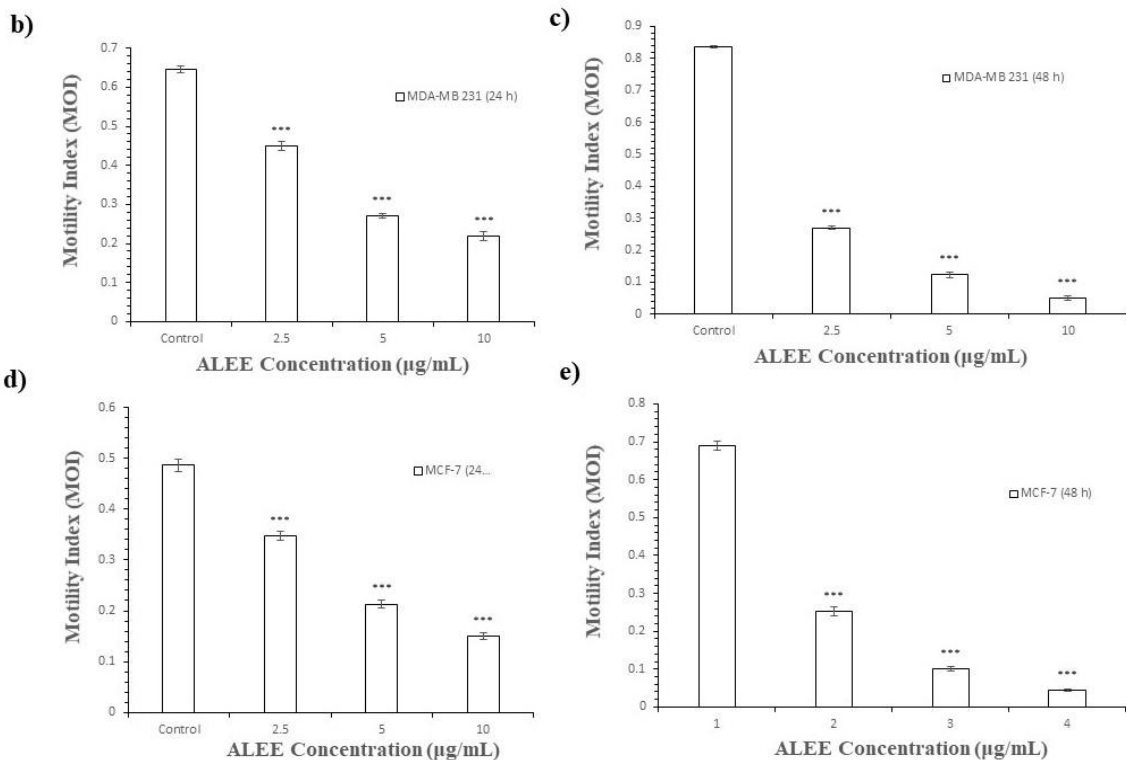
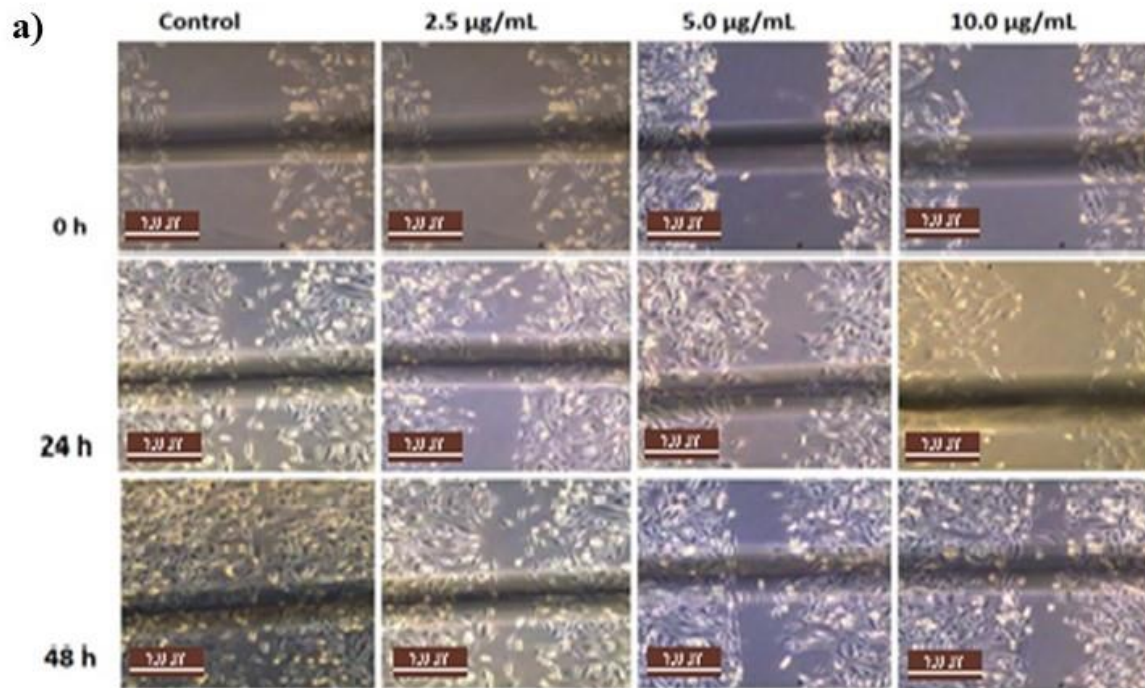


Figure 6. Effect of Quercetin on Lateral Motility of MDA-MB 231 and MCF. Data are presented as mean $\pm$ SEM of at least replicates experiments ($n \geq 3$).

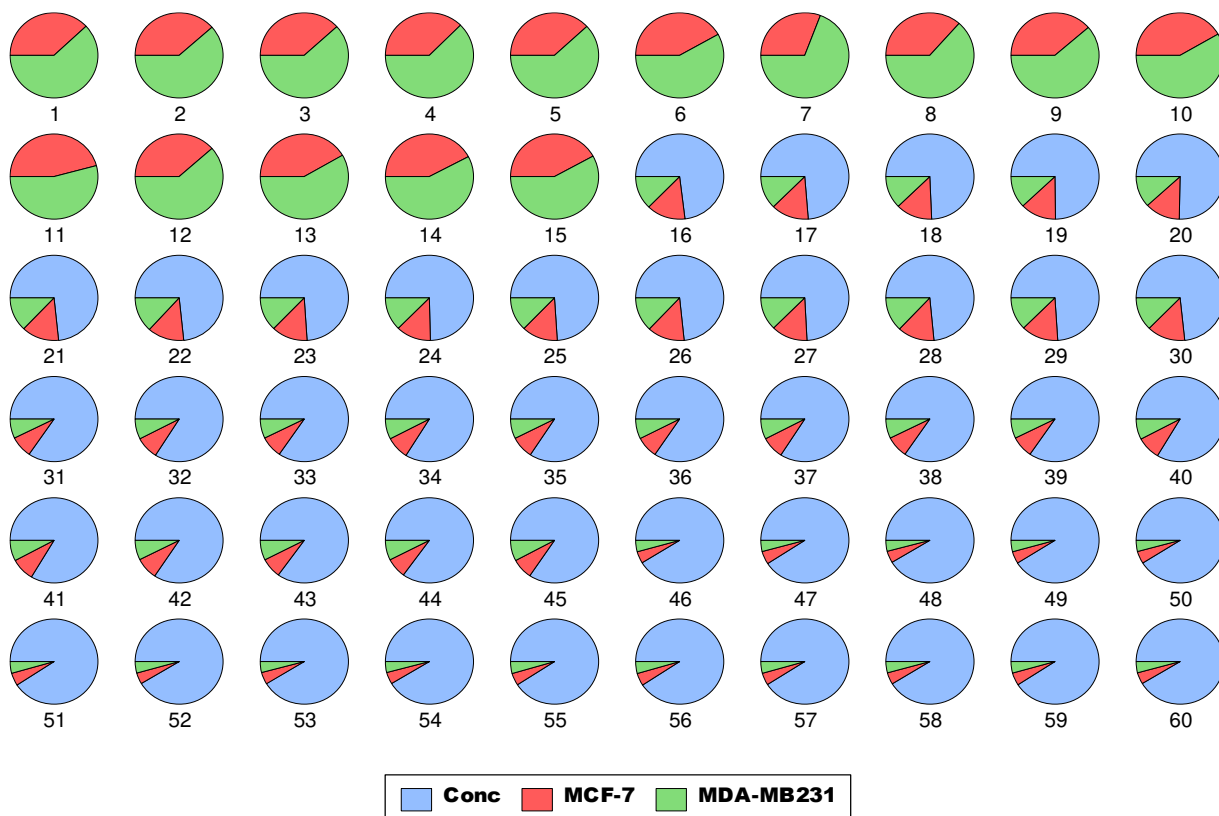


Figure 7. Visualized pie-chart distribution

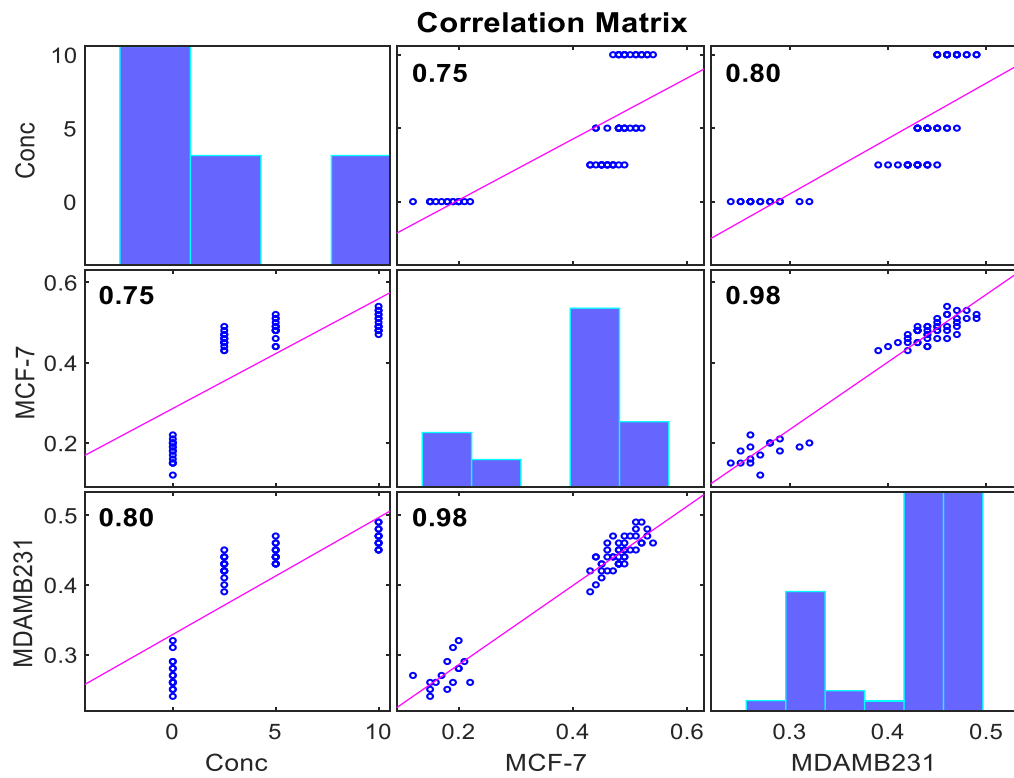
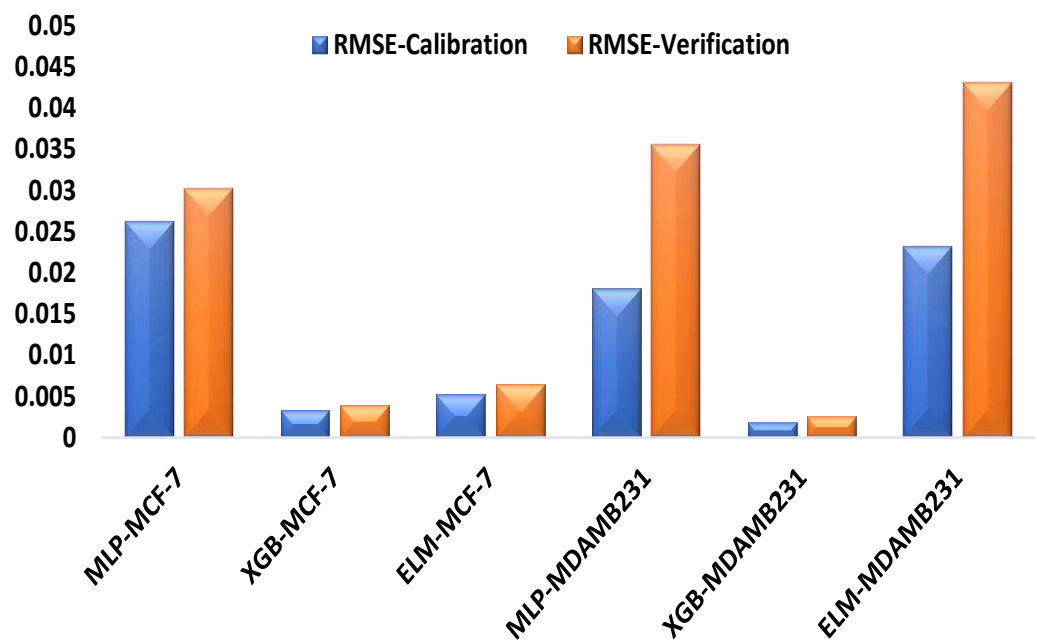


Figure 8. Correlation matrix between the raw data

719



720

721 **Figure 9.** Relative mean square error (RMSE) for predicting of MDA-MB 231 and MCF-7
722 human Breast Cancer treated with ALEE in both the calibration and verification stages.

723

724

725

726

727

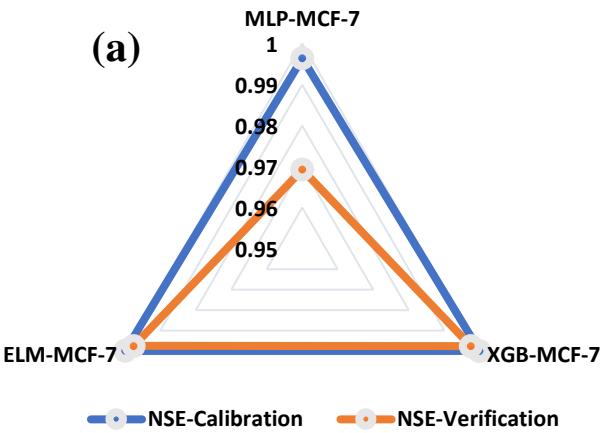
728

729

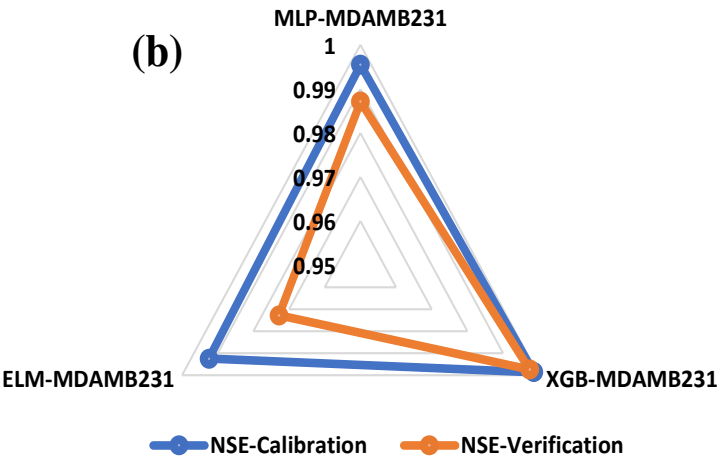
730

731

732



733



734

735 **Figure 10.** Radar chart for various variation determination and correlation coefficients (a)
736 MCF-7 (b) MDA-MB 231.

737