

PAPER • OPEN ACCESS

Development of predictive models for cervical cancer based on gene expression profiling data

To cite this article: A A Abdullah *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **557** 012003

View the [article online](#) for updates and enhancements.

Development of predictive models for cervical cancer based on gene expression profiling data

A A Abdullah¹, N K Abu Sabri¹, Wan Khairunizam¹, I Zunaidi², Z M Razlan¹ and Shahrman A B¹

¹School of Mechatronic Engineering, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia

²Faculty of Technology, University of Sunderland, St Peter's Campus, Sunderland, SR6 0DD, United Kingdom

Abstract. Cervical cancer and the prediction of clinical outcome are among the most important emerging applications of gene expression microarray technology with feature sequencing of microRNA. By using reliable and dependable classification of machine learning algorithms available for microarray gene expression profiling data is the key in order to develop the most suitable and possible predictive model to be used by patient. In this paper, two-machine learning algorithms have been used which are Support Vector Machine (SVM) and Random Forests (RF) for the predictive models of cervical cancer. We identify and evaluate the performance of these two algorithms in order to know which algorithm has better performance. In this study, 714 features and 58 samples are used to develop predictive model for cervical cancer and our computational results show that RF algorithm outperform SVM algorithm with the accuracy of 94.21%. Our data also underline the importance of variables, which give the significant role in predicting the occurrence of cervical cancer.

1. Introduction

Cervical cancer is the second most common cancer among women. It may cause mortality and morbidity [1], [2]. The high mortality rate of the cervical cancer is because of the lack of awareness of women for early detection of cervical cancer [3]. Even though cervical cancer is dangerous which may lead to life threatening but it is potentially curable. Cervical cancer occurs in a woman's cervix [4]. The cervix is the lower part of the uterus. It connects the uterus to the vagina [4]. There are two main types of cervical cancer, which are squamous cell carcinoma and adenocarcinoma. The squamous cell carcinoma cancer type usually occurred in the epithelium lining of the cervix [5]. As for adenocarcinoma cancer type, it is developed from gland cells. The adenocarcinoma of cervical cancer is a premalignant glandular condition [6]. Squamous cell carcinoma has the highest percentage of cervical cancer cases which are 90% and 10% of cervical cancer cases are adenocarcinoma [1]. Hence, adenocarcinoma neoplasia of the cervical cancer is less common than squamous cell carcinoma neoplasia of cervical cancer when it is being diagnosed [6].

Cervical cancer is caused by a virus named human papillomavirus (HPV). When the infection of the human papillomavirus (HPV) at the cervix left untreated, cervical cancer is developed [7]. In cervical cancer, human papillomavirus (HPV) is the most important infectious agent because it contributes to neoplastic progression. Neoplastic progression is the progression of the abnormal growth of the cervical cancerous cells and proliferation of the abnormal cells due to a malignant process [8]. However, most scientific studies have found that human papillomavirus (HPV) infection alone is inadequate to induce the malignant of cervical cancer. Other host genetic variations also play



important role in the development of cervical cancer in women [1]. Commonly, microarray gene expression profiling is used to identify and differentiate the expressed gene expression in a pre-cancerous and cancerous cell in the cervix [1]. A microarray or also known as DNA chip is an arrangement of DNA molecules that have been chemically bonded to a fine grid of surfaces. The purpose of the microarray gene expression is to interpret and analyze the genes expression state in complementary DNA prepared from mRNA in which the hybridization is taking place on the array [9]–[11]. Genome-wide expression profiling has the potential to get more accurately predict outcome at the end of the research since it allows selection of the genes that associate most strongly with the outcome by screening a large number of these genes simultaneously. Expression profiling studies show that genes are statistically significant differences under changed experimental conditions [12].

In this study, the predictive model for determining whether the cells are cancerous or pre-cancerous for cervical cancer can be developed by just looking at their gene expression profiles. In order to develop the predictive model, machine learning algorithms approaches have been implemented. Artificial intelligence and machine learning techniques have been applied in various medical diagnostic systems [13]–[16] and there are many researchers utilized the techniques in the image processing of Pap smear images [17]–[23]. However, there are fewer studies on implementing machine learning methods on the cervical cancer based on gene expression profiling data. Hence, in order to overcome this problem, we implemented unsupervised and supervised machine learning techniques, which are Principal Components Analysis (PCA), Support Vector Machine (SVM) and Random Forest (RF), respectively. The Principal Components Analysis (PCA) is used to do the data screening and data preprocessing for the model. The data that have been preliminaries in the Principal Components Analysis (PCA) will be used in Support Vector Machine (SVM). As for support machine vector (SVM) and Random Forest (RF), they are used to predict the accuracy of the predictive model for cervical cancer whether the cell is pre-cancerous or cancerous cells.

2. Proposed Methods

In this study, we used dataset, which is obtained from the Gynecologic Oncology Group Tissue Bank (PA, USA) for classification of cervical cancer based on the gene expression profiling data [24]. First, the dataset will be undergoing data extraction process in order to get and retrieving the relevant data or information in the dataset. The data pre-processing then applied on the dataset for eliminating the irrelevant and redundant data containing in the dataset. In order to perform classification of the gene expression profiling data of microarray dataset into their cluster, a tree-like structure is constructed by using hierarchical clustering. By implemented machine learning algorithms of Support Vector Machine (SVM) and Random Forests (RF) on the dataset, predictive model for cervical cancer can be developed. In Support Vector Machine (SVM) model, Principal Components Analysis (PCA) algorithm is used as for the feature selection technique. It is very useful in order to reduce the high dimensionality of dataset as to facilitate the algorithm to produce a good performance. The new dataset constructed will be executed by using Support Vector Machine (SVM) model in order to get the prediction accuracy. As a comparison, Random Forests (RF) is also applied to get the accuracy of predictive model for cervical cancer.

2.1. Data extraction

Data extraction is a process of retrieving relevant data or information out of data sources, which is dataset containing microarray data analysis. In this stage, the data will be extracted and loaded into the staging area of the relational database of the cervical cancer. When data extraction is implemented, the extraction logic is applied in order to extract the data from the dataset. The data column of the dataset is the data regarding the normal and tumor of samples data. And as for the data in the row is about the features of the sequencing of microRNAs. The data matrix produce from the data extraction is changed into a data frame because it may ease us to interpret the data for further processing.

2.2. Data pre-processing

Data pre-processing is an important step in the data mining process. It is implemented on the dataset because the dataset obtained from microarray analysis contains irrelevant, unreliable and redundant of data or noise present in the dataset. Data pre-processing is performed by undergoing the process of data cleaning, data scaling and normalization of data. In data cleaning process, the corrupt or irrelevant data from the datasets are detected. Then, the data that have been identified as incomplete and irrelevant will be filtered out of the dataset. After cleaning the corrupted data, a final dataset should be consistent with other similar dataset in the system and can be used for further processing. Other than data cleaning process, data scaling is also implemented on the dataset to pre-processing the dataset. In data scaling, the data will be centering and scaling whenever there are data with different units and ranges. By scaling the data, the final dataset will be easy to be interpreted. The mean value of sample data from the datasets is reduced by using the centering transformation method. As for scaling transformation, it will divide the value of each sample data by the standard deviation of the dataset. Thus, the numerical data of the dataset stability is improved. The normalization of the data is important to organize the columns and tables of the database to reduce the redundancy of the data. It is used to improve the integrity of the dataset in this study.

2.3. Hierarchical clustering

The subgroups are defined based on the genes expression profiling data of the dataset. In order to classify the gene expression profiles of data into their own clusters, hierarchical clustering is implemented to create the hierarchical, which is the tree-like structure of the data. It is also referred as a dendrogram. The hierarchical clustering is implemented by measuring the feature similarities of the gene expression profiles. Hierarchical clustering is implemented by measuring the maximum distance matrix between the samples data. In every stage of hierarchical clustering, the two nearest clusters are merged together into a new cluster or subtypes. The clustering process is repeated until all of the samples data is agglomerated into a single cluster.

2.4. Principle component analysis (PCA)

In order to develop the predictive model by using the support vector machines (SVM), principal components analysis (PCA) is used as dimensional reduction technique because it is useful to collapse the features into a smaller set of principal components. The principal components analysis (PCA) is used in this study because the dataset contains hundreds of features from the sequencing of microRNA. The principal components analysis (PCA) is needed in order to reduce the dimensionality of the hundreds of features in the dataset. By implemented the machine learning algorithm on the dataset, it helps to collapse the hundreds of features into a smaller set of principal components.

2.5. Support vector machine (SVM)

Support vector machine (SVM) algorithm is implemented by performing a classification on the final dataset. The classification of the algorithm is implemented by constructing a multidimensional hyperplane to the selective samples data. It will maximize the margin between the two-data clusters, which are normal (N) and tumor (T). The type of the SVM parameter used in this study is set to the C-classification. As for the kernel, linear of kernel type is selected.

In order to evaluate the predictive model performance, classification of accuracy, sensitivity and specificity need to be calculated, as shown in (1), (2) and (3).

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \quad (1)$$

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (3)$$

TN is the number of true negative, FP is defined as the number of false positive, TP is defined as number of true positive and FN is the number of false negative. These coefficients are defined and illustrated as a confusion matrix in the Table 1.

Table 1. Confusion matrix.

Prediction	Reference	
	Negative (0)	Positive (1)
	Negative (0)	Positive (1)
	TN	FP
	FN	TP

2.6. Random forest (RF)

In order to make sure the result obtained is valid and can be generated for generating the prediction model, the dataset is randomly partitioned into training and testing sets. In this project, 70% of the samples data are used for training and the rest of the 30% for testing. The type of the RF parameter is set to the classification. As for the number of trees in the RF (ntree), it is set to 300. ntree is the number of trees to grow in the Random Forest (RF). The ntree should not be set to a small number because we need to ensure that every input row gets predicted at least a few times. The last RF parameter that needs to be considered is number of variables tried at each split (mtry). mtry is a number of variables that randomly sampled as candidates at each split. The number of variables tried at each split can be calculated by using (4), where p is number of variables in dataset. After the calculation is done, it is set to 27.

$$\sqrt{p} = \sqrt{714} = 26.72 \approx 27 \quad (4)$$

In order to evaluate the accuracy of random forest algorithm, OOB estimate of error rate from RF module and predictive model performance by calculating accuracy, sensitivity and specificity are considered. For both model performances, confusion matrix as illustrated in Table 1 is constructed.

3. Results and Discussion

3.1. Hierarchical clustering

The result of the clustering is presented using a heat map, as shown in Figure 1. Hierarchical clustering is mapped in the heatmap is using one-channel microarray data, which contain the gene expression profiles data. As we can see from Fig. 1, it contains several colors with two dendrogram. The tree-like structures presented above and the left of the heatmap are known as dendrogram. It works by calculating the pairwise distance between all of the data in the dataset. The data with least distant apart will be merged together until rest of the data have been agglomerated to represent as the dendrogram as illustrated in Figure 1. The samples data, which contain normal and tumor, are shown as rows by one row per data sample. The columns represent the features of the sequencing microRNAs below the heatmap figure. The features of sequencing microRNA levels are represented by the color scale which ranging from orange to white according to heat color. Orange indicates a low level of features of sequencing microRNA level while white indicates high level of features of sequencing microRNA level. The tree-like structure (dendrograms also provide some qualitative means of assessing the similarity between the features of microRNA and data samples.

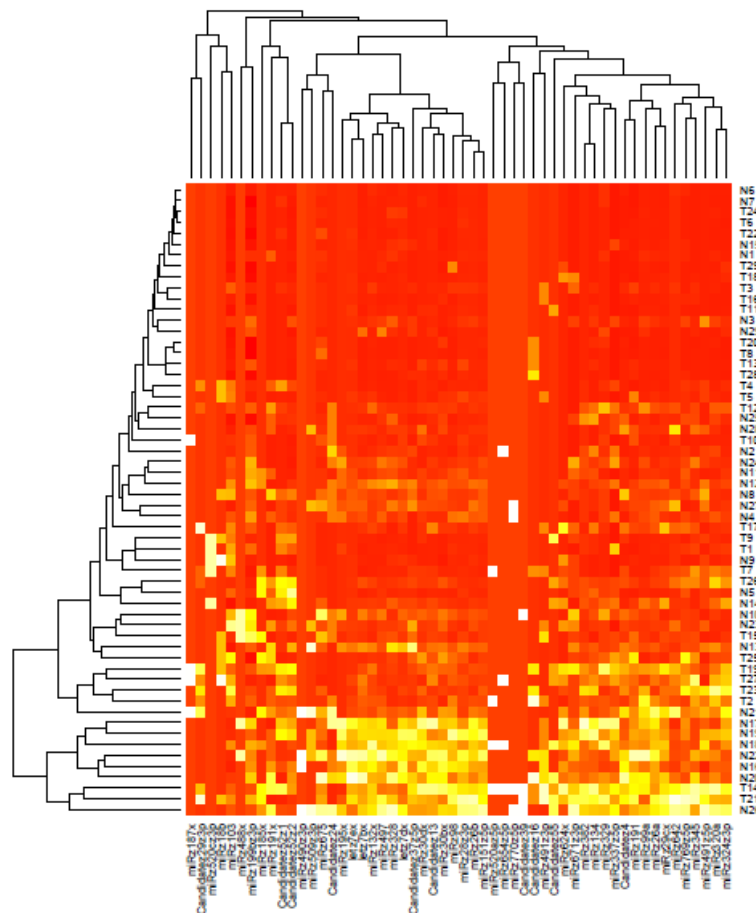


Figure 1. Hierarchical clustering by using heatmap.

3.2. Support vector machine (SVM)

According to the test data, there are 16 sample data that have been used for testing in order to develop the predictive model are shown in Table 2. In order to perform the prediction on the sample data, confusion matrix needs to be evaluated. The confusion matrix is demonstrated in the Table 3. Based on the confusion matrix in Table 3, we can see the performance of the support vector machine (SVM) algorithm for prediction. For the normal sample data, the model has predicted that 7 out of 8 sample data are normal while the rest of the sample data, which is only 1 sample data, has been predicted as tumor. As for the tumor sample data, 7 out of 8 samples have been predicted as tumor by the model and only 1 sample data has been predicted as normal. From the confusion matrix, the performance of support vector machine (SVM) algorithm model can be evaluated and demonstrated in Figure 2.

Table 2. Selective samples data for SVM.

	Normal	Tumor
Selective Samples Data	N12	T2
	N17	T4
	N18	T9
	N20	T10
	N24	T12
	N25	T13
	N27	T14
	N29	T22

Table 3. Confusion matrix for SVM (1 indicates as tumor while 0 indicates as normal).

		Reference	
Prediction		Negative (0)	Positive (1)
	Negative (0)	7	1
	Positive (1)	1	7

```

Accuracy : 0.875
95% CI : (0.6165, 0.9845)
No Information Rate : 0.5
P-value [Acc > NIR] : 0.00209

Kappa : 0.75
McNemar's Test P-Value : 1.00000

Sensitivity : 0.8750
Specificity : 0.8750
Pos Pred Value : 0.8750
Neg Pred Value : 0.8750
Prevalence : 0.5000
Detection Rate : 0.4375
Detection Prevalence : 0.5000
Balanced Accuracy : 0.8750

'Positive' class : 0

```

Figure 2. Prediction result for cervical cancer using SVM algorithm.

Based on Figure 2, the accuracy of the predictive model is 0.8750. The Kappa statistic is calculated for the overall accuracy of the predictive model. Kappa statistic value of the predictive model is 0.75, which indicates a good arrangement. The value for both sensitivity and specificity of the predictive model are 0.8750. Hence, we can conclude that the developed predictive model of cervical cancer based on the gene expression profiles data is a good prediction. Since, the accuracy of the predictive model is up to the 80%. So, it is proved that SVM algorithm can be used to develop the predictive model of cervical cancer based on the gene expression profiling data.

3.3. Random forest (RF)

The reliability and accuracy of the RF prediction model performance for cervical cancer of gene expression profiling data is shown in the Figure 3. According to the training data used in this study, there are 41 sample data that have been selected in order to train the data. Thus, the accuracy of predictive model produced at the end of this study is high which is greater than 80% as desired outcome. The testing data used throughout the process is shown in Table 4. For RF algorithm, cross-

validation is not required in order to get the unbiased estimate of the test set error. In training data stage of Random Forests (RF) model, out-of-bag (OOB) estimate of error rate is determined. OOB is the error of mean prediction on each training sample. In order to estimate the prediction error, the data used in the estimation is not in the bootstrap sample for each bootstrap iteration and related tree. Generally, OOB error rate is estimated internally during the program run. The OOB estimate of error rate and the confusion matrix for the RF model is demonstrated in Figure 3.

Table 4. Selective samples data for RF.

	Normal	Tumor
Selective Samples Data	N2	T2
	N4	T3
	N5	T5
	N8	T8
	N11	T21
	N16	T24
	N20	T29
	N21	
	N24	
	N26	

```

Type of random forest: classification
Number of trees: 300
No. of variables tried at each split: 27

OOB estimate of error rate: 14.63%
Confusion matrix:
 0  1 class.error
0 15  4  0.21052632
1  2 20  0.09090909

```

Figure 3. Out-of-bag (OOB) estimate of error rate and confusion matrix in RF model.

As shown in the Figure 3, out-of-bag (OOB) estimate of error rate in RF model is 14.63%. This means that the accuracy of the RF model is quite good because it is up to 85%. In order to assess the performance of predictive model by using Random Forest (RF) algorithm for cervical cancer, confusion matrix needs to be evaluated by using testing data. The confusion matrix for RF is demonstrated in Table 5.

Table 5. Confusion matrix for RF (1 indicates as tumor while 0 indicates as normal).

		Reference	
Prediction		Negative (0)	Positive (1)
	Negative (0)	10	1
	Positive (1)	0	6

According to the confusion matrix in Table 5, the visualization of prediction performance by using RF algorithms can be analyzed. The model has predicted that 10 out of 10 are the normal sample data. None of normal samples is predicted as tumor. As for the tumor samples data, 6 out of 7 samples data are predicted as tumor and only 1 sample data is predicted as normal. From the confusion matrix in Table 5, the performance of prediction by Random Forests (RF) algorithms can be demonstrated in Figure 4.


```

Accuracy : 0.9412
95% CI : (0.7131, 0.9985)
No Information Rate : 0.5882
P-value [Acc > NIR] : 0.001559

Kappa : 0.8759
McNemar's Test P-value : 1.000000

Sensitivity : 1.0000
Specificity : 0.8571
Pos Pred Value : 0.9091
Neg Pred Value : 1.0000
Prevalence : 0.5882
Detection Rate : 0.5882
Detection Prevalence : 0.6471
Balanced Accuracy : 0.9286

'Positive' class : 0

```

Figure 4. Prediction result for cervical cancer using RF algorithm.

As illustrated in the Figure 4, the accuracy of the predictive model is 0.9412, which in percentage is approximately 94.12%. This prediction accuracy by using RF algorithm is higher than SVM model. Besides that, kappa statistic value of the predictive model is 0.8759, which indicates a substantial arrangement. The sensitivity of this predictive model is perfect which 1.0. This can be seen in the confusion matrix in Table 5 where there is no error in predicting the normal sample data. On the other hand, the specificity of predictive model is 0.8571, which show that it is a good predictive model. In RF model, variables that play importance role in interpretation and prediction can be evaluated. The variables importance with Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG) is shown in Figure 5.

Top 10 - Variable Importance

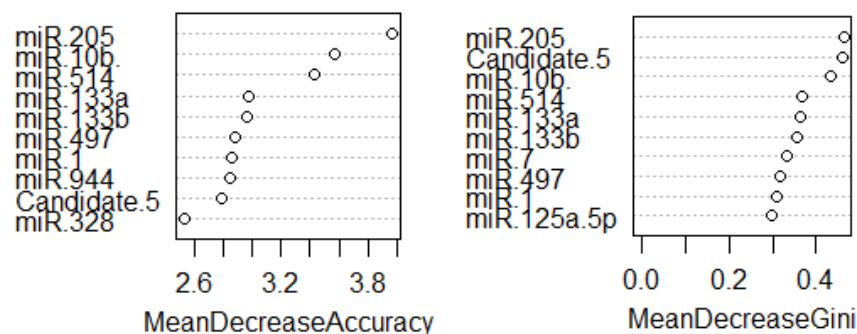


Figure 5. The top 10 of variables importance in RF model.

As illustrated in the Figure 5, each variable plotted in the graph indicates how important the variables are in the classification of RF model. On the x-axis, the importance of the variables is plotted and the variables importance involved in the classification on y-axis. The variables are plotted by plotting the most important variables at the top to the least important variables by making a dot-plot on x-axis. Based on Figure 5, miR. 205 candidates is the most important variable for both MDA and MDG with the value of 3.96919433 and 0.462774856 respectively. According to the Zhang Yue et al, they stated that miR. 205 is a tumor-suppressing onco-miRNA [25]. In addition, Hong Xie et al. have concluded that miR. 205 plays an important role in pathogenesis of cervical cancer by promoting cell proliferation and migration [26]. miR. 10b is also one of the variables importance in predicting cervical cancer with the value of 3.57406978 for MDA and 0.434711805 for MDG. Besides that, in

Zhang Yue et al. studies, they concluded that miR. 10b is able to promote the cervical cancer tumor invasion [25]. Ren Hou et al. reported that miR. 10b is one of the tumor suppressor microRNA which downregulated in cancer tissues. But it will inhibit cervical cancer growth and metastasis when it is in overexpression state [27]. Dongling Zou et al. stated that the anomalous expression of miR. 10b in cervical cancer shows its play role in determining cervical carcinogenesis [28]. Thus, depending on the data recorded as shown in Figure 5, we are able to predict that the sample data of gene expression profile with a high value of this top 10 of the important variables are tumor. The investigation of miRNA levels in cervical cancer cells certainly can open new possibilities for studying molecular markers in the context of early diagnosis and prognosis programs [29].

4. Conclusion

In this study, we used dataset, which is obtained from the Gynecologic Oncology Group Tissue Bank (PA, USA) [24] for classification of cervical cancer based on the gene expression profiling data. We conducted the comprehensive benchmarking of Support Vector Machine (SVM) and Random Forests (RF) to evaluate the performance of the predictive models based on the accuracy, Kappa value, sensitivity and specificity. Based on our computational result, we can conclude and prove that Random Forests (RF) machine learning algorithm can be successfully used for predicting cervical cancer based on the gene expression profiling data with the microarray dataset. The accuracy obtained in cervical cancer prediction is better than previous researches results. The average Random Forests (RF) model's accuracy obtained is 94.21%, which may be acceptable in many applications. Besides that, the investigation for a high overall performance classifier with microarray gene expression and other "omics" data is still in progress. Random Forests (RF) algorithms exhibit one of the best classification performances in predicting cervical cancer.

References

- [1] C. M. Martin, K. Astbury, L. McEvoy, S. O'Toole, O. Sheils, and J. J. O'Leary, "Gene expression profiling in cervical cancer: identification of novel markers for disease diagnosis and therapy.," *Methods Mol. Biol.*, vol. 511, pp. 333–359, 2009.
- [2] M. Lupu, C. Caruntu, M. A. Ghita, V. Voiculescu, S. Voiculescu, A. E. Rosca, A. Caruntu, L. Moraru, I. M. Popa, B. Calenic, M. Greabu, and D. E. Costea, "Gene Expression and Proteome Analysis as Sources of Biomarkers in Basal Cell Carcinoma," *Disease Markers*, vol. 2016, 2016.
- [3] S. W. Purnami, P. M. Khasanah, S. H. Sumartini, V. Chosuvivatwong, and H. Sriplung, "Cervical cancer survival prediction using hybrid of SMOTE, CART and smooth support vector machine," in *AIP Conference Proceedings*, 2016, vol. 1723.
- [4] D. R. Tobergte and S. Curtis, *Principles of Biochemistry 5th ed*, vol. 53, no. 9, 2013.
- [5] M. Volpacchio, J. C. Vilanova, and A. Luna, "Cervix and vagina," *Learn. Imaging*, vol. 1, no. 6, pp. 165–186, 2012.
- [6] M. P. Hopkins, J. a Roberts, and R. W. Schmidt, "Cervical adenocarcinoma in situ.," *Obstetrics and gynecology*, vol. 71, no. 6 Pt 1, pp. 842–844, 1988.
- [7] A. Zayed, "Cervical Cancer: Types, Symptoms, Causes, Treatments and Medications," *ConsumerHealthDigest*, 2017. [Online]. Available: <https://www.consumerhealthdigest.com/health-conditions/cervical-cancer.html>.
- [8] A. Deverakonda and N. Gupta, "Diagnosis and treatment of cervical cancer: a review," *J. Nurs. Heal. Sci.*, vol. 2, no. 3, pp. 1–11, 2016.
- [9] "DNA Microarray microarray," in *Encyclopedia of Public Health*, 2008, pp. 298–299.
- [10] M. Debnath, G. B. K. S. Prasad, and P. S. Bisen, "Microarray," in *Molecular Diagnostics: Promises and Possibilities*, 2010, pp. 193–208.
- [11] A. M. Charpe, "DNA microarray," in *Advances in Biotechnology*, vol. 9788132215547, 2013, pp. 71–104.
- [12] A. H. Klopp and P. J. Eifel, "Gene expression profiling in cervical cancer: State of the art and

- future directions,” *Cancer Journal*, vol. 12, no. 3. pp. 170–174, 2006.
- [13] A. A. Abdullah, B. S. Chize, and Y. Nishio, “Implementation of an improved cellular neural network algorithm for brain tumor detection,” in *2012 International Conference on Biomedical Engineering (ICoBE)*, 2012, pp. 611–615.
 - [14] A. A. Abdullah and S. Kanaya, “Prediction of Biological Activities of Volatile Metabolites Using Molecular Fingerprints and Machine Learning Methods,” *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 1–17, pp. 91–96, 2018.
 - [15] A. A. Abdullah, N. S. Fadil, and W. Khairunizam, “Development of Fuzzy Expert System for Diagnosis of Diabetes,” in *2018 International Conference on Computational Approach in Smart Systems Design and Applications, ICASSDA 2018*, 2018.
 - [16] A. A. Abdullah, A. Yaakob, and Z. Ibrahim, “Prediction of Spinal Abnormalities Using Machine Learning Techniques,” in *2018 International Conference on Computational Approach in Smart Systems Design and Applications, ICASSDA 2018*, 2018.
 - [17] A. A. Abdullah, A. Fonetta, D. Giong, N. Adilah, and H. Zahri, “Cervical cancer detection method using an improved cellular neural network (CNN) algorithm,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 14, no. 1, pp. 210–218, 2019.
 - [18] N. A. Obukhova, A. A. Motyko, U. Kang, S.-J. Bae, and D.-S. Lee, “Automated image analysis in multispectral system for cervical cancer diagnostic,” *Conf. Open Innov. Assoc. Fruct*, vol. 2017–April, 2017.
 - [19] P. Sukumar and R. K. Gnanamurthy, “Computer Aided Detection of Cervical Cancer Using Pap Smear Images Based on Adaptive Neuro Fuzzy Inference System Classifier,” *J. Med. Imaging Heal. Informatics*, vol. 6, no. 2, pp. 312–319, 2016.
 - [20] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, “A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images,” *Computer Methods and Programs in Biomedicine*, vol. 164, pp. 15–22, 2018.
 - [21] M. Orozco-Monteagudo, A. Taboada-Crispi, and H. Sahli, “Biologically inspired anomaly detection in pap-smear images,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 8259 LNCS, no. PART 2, pp. 17–24.
 - [22] B. Taha, J. Dias, and N. Werghi, “Classification of cervical-cancer using pap-smear images: A convolutional neural network approach,” in *Communications in Computer and Information Science*, 2017, vol. 723, pp. 261–272.
 - [23] Abid Sarwar, “Analysis of Machine Learning and Statistics Tool Box (Matlab R2016) over Novel Benchmark Cervical Cancer Database,” *Int. J. Trend Sci. Res. Dev.*, vol. 2, no. 1, pp. 619–622, 2018.
 - [24] D. Witten, R. Tibshirani, S. G. Gu, A. Fire, and W. O. Lui, “Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls,” *BMC Biol.*, vol. 8, 2010.
 - [25] Z. Yue, Z. Yun-Shan, and X. Feng-Xia, “miR-205 mediates the inhibition of cervical cancer cell proliferation using olmesartan,” *JRAAS - J. Renin-Angiotensin-Aldosterone Syst.*, vol. 17, no. 3, 2016.
 - [26] H. Xie, Y. Zhao, S. Caramuta, C. Larsson, and W. O. Lui, “miR-205 Expression Promotes Cell Proliferation and Migration of Human Cervical Cancer Cells,” *PLoS One*, vol. 7, no. 10, 2012.
 - [27] R. Hou, D. Wang, and J. Lu, “MicroRNA-10b inhibits proliferation, migration and invasion in cervical cancer cells via direct targeting of insulin-like growth factor-1 receptor,” *Oncol. Lett.*, vol. 13, no. 6, pp. 5009–5015, 2017.
 - [28] D. Zou, Q. Zhou, D. Wang, L. Guan, L. Yuan, and S. Li, “The Downregulation of MicroRNA-10b and its Role in Cervical Cancer,” *Oncol. Res. Featur. Preclin. Clin. Cancer Ther.*, vol. 24, no. 2, pp. 99–108, 2016.
 - [29] B. Pardini, D. De Maria, A. Francavilla, C. Di Gaetano, G. Ronco, and A. Naccarati, “MicroRNAs as markers of progression in cervical cancer: A systematic review,” *BMC Cancer*, vol. 18, no. 1, 2018.