ISSN 2278-3091
Volume 10, No.2, March - April 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse861022021.pdf
https://doi.org/10.30534/ijatcse/2021/861022021

Cell Cycle-Regulated Genes Classification using Machine Learning and Deep Learning Techniques on Processed Microarrays Images

Hiba Lahmer[1], AfefElloumi Oueslati[1,2], Zied Lachiri[1]
[1]University of Tunis El Manar, SITI Laboratory National School of Engineers of Tunis, Tunisia,
lahmer.hiba@gmail.com
[2]University of Carthage, National School of Engineers of Carthage,Electrical Engineering Department, Tunisia,
Afef.Elloumi@enit.utm.tn
[1]University of Tunis El Manar, SITI Laboratory National School of Engineers of Tunis,Tunisia,
Zied.lachiri@enit.rnu.tn

ABSTRACT

Nowadays, machine learning and deep learning algorithms, are considered as new technologies increasingly used in the biomedical field. Machine learning is a branch of Artificial Intelligence that aims to automatically find patterns in existing data. A new Machine Learning subfield, the deep learning theory, has emerged. It deals with object recognition in images. In this paper, our goal is DNA Microarrays'analysis with these algorithms to classify two genes' types. The first class represents cell cycle regulated genes and the second is non cell cycle regulated ones. In the current state of the art, the researchers are processing the numerical data associated to gene evolution to achieve this classification. Here, we propose a new and different approach, based on the microarrays images' treatment. To classify images, we use three machine learning algorithms which are: Support Vector Machine, K-Nearest Neighbors and Random Forest Classifier. We also use the Convolutional Neural Network and the fully connected neural network algorithms. Experiments demonstrate that our approaches outperform the state of art by a margin of 14.73 per cent by using machine learning algorithms and a margin of 22.39 per cent by using deep learning models. Our models accomplish real time test accuracy of ~ 92.39 % at classifying using CNNand 94.73% using machine learning algorithms.

Keywords: DNA microarrays, cell cycle, regulated gene, machine learning, classification, SVM, KNN, RFC, deep learning, CNN,fully connected neural network

1.INTRODUCTION

The Artificial Intelligence causes a big revolution in the biomedical industry. We are treating now an important quantity of data. To find and extract knowledge, we must apply the adequate tools and techniques. So, we need to train the machine, to become able to imitate human intelligence [1].Gene expression microarrays increasingly promise for analysis, diagnosis and prediction of clinical decisions. In order to exploit this technology, scientists are repeatedly searching to apply the classification algorithms for the gene characteristic's creation for each class [2]. Supervised learning, find a function and define the boundaries that maps an input to an output, based on a training set of examples input-output [3]. We use this function to predict the classification used for unlabeled data. So many applications and tools, make inroads toward using and exploring big biomedical data. DNA microarray is one of the fastest-growing technologies that generates high-dimensional molecular data [4].

This technology is used to measure the mRNA transcription levels. It permits the simultaneous visualization of the expression levels of thousand genes [5]. It is provided for a specific pathological context or a cell type. It permits the detection of changes in the nucleotide sequence, so we can identify the similarity in gene groups, having a particular pattern [6]. For example, to classify a cancer, we must select and recognize the genes linked to this cancer, by analyzing the microarrays images and exploring various functionalities and classifiers [7]. Many studies have identified many types of cancer by using the DNA Microarrays data and applying machine learning algorithms and deep learning models [8]. Other use of this technology permits the identification of the cell cycle regulated genes and extract changes in genes' behavior. Many studies turn around the regulation of the cell cycle in several organisms, giving a wide range of high-throughput technologies, as DNA microarrays [9].

In eukaryotes,the cell cycle expresses periodically, it is a regulated process that is composed of events and involves thousands of proteins. The cell cycle regulated genes must be ingrained during the progression, that's why they are cyclic expressed. These genes are the orchestra of the cycle's regulation [10].This process is divided into four phases. The first stage is “S” where the DNA replication takes place. The second stage is “M” where the chromosomes stand out from each other. The two other phases of the cycle, “G1” and “G2”, represent intervals (Gap). During the G1 phase, the cell grows, integrates the mitogenic or anti-mitogenic signals and prepares to correctly perform the S phases. When during phase G2, the cell is preparing for phase M [11].

In order to analyze and understand this process, we must focus on the behavior of the cell cycle regulated genes. Many methods have been proposed; the first large scale identification of these genes is made in the genome of Saccharomyces Cerevisiae with classical methods in1991. They discovered that there is a set of genes regulated periodically with the cell cycle. Therefore ,many of them are involved in regulating this cycle. Spellman et al [12] proposed a comprehensive series of experiments to recognize all protein-encoding transcripts in the yeast genome that are cell cycle-regulated. Then, they used the DNA microarrays to supervise mRNA levels changes during the cell cultures over three methods. After, they analyzed the data by deriving a score based on Fourier calculation and a correlation function that detect genes regulated by the cell cycle using correlation and periodicity [12]. They identified 800 genes that accord to an objective minimum criterion for cell cycle regulated genes. Using this database and analysis, Chenglin Liu, Peng liu and Tow Hug developed a framework for cell cycle gene identification based on the Convolutional Neural Network. Their approach uses three categories, representing stable, increase and decrease of gene expression changes [14]. Their approach consists on the conversion of the expression data to categorical state with values -1, 0 and 1, respectively attributed to the three categories [13]. They tried six machine learning algorithms such as KNN and SVM. To identify cell cycle regulated genes using deep learning, different approaches were tried. One of them, is an algorithm named DLGene. DLGene is based on CNN. It consists of an input layer, one convolutional layer, two fully connected neural network layers and an output layer. The input layer takes 4*14 state matrix profiling the expression changes of one gene [13].

In this paper, we propose a new way for processing and analyzing DNA Microarray to recognize the cell cycle regulated genes using machine learning techniques and deep learning models. It is composed of five sections. The second section presents the machine learning techniques. In the section three, we present the deep learning methods. Section four explains the proposed method.In section five, we discuss the experimental results. The last section gives the conclusion of this paper.

2. Machine learning techniques

The amount of biomedical data is growing exponentially. Our goal is to explore and to analyze this data by using the suitable tools and techniques. So, we need to apply Machine Learning to find the hidden patterns and extract knowledges. It is a set of methods and techniques that permit the machine training to imitate human intelligence and to make easier a decision. There are four types of learning. The first type is the supervised learning, used to predict the discrete category of unlabeled data, called classification. we are talking about a regression, if the output of prediction consists on continuous variables. The classification of TSSs is an example of use of supervised learning on biological domain is [14]. The second type is the non-supervised learning. It is a set of algorithms discovering the hidden data structure by looking for patterns and similarities between groups, based on the distance calculation. The number of clusters and their nature have not been predefined previously for example the exon and intron regions [15]. The third type, the semi-supervised learning, combines the supervised and unsupervised learning so we recognize object from a small amount of labeled data and a large amount of unlabeled data during training. The labeled data is used to identify that there are specific groups or types in the input data and what they might be. Finally, the reinforcement learning, is alongside supervised learning and unsupervised learning. In the absence of training dataset, the reinforcement agent decides sequentially what to do to perform a given task. It has to learn from its experience, so we give labels to sequences of dependent decisions [16].

In machine learning, classification is a supervised learning approach. It learns from the data input given to it, discover a set of new information and then uses it to classify new observation. It proposes to classify new data in line to a deduced model from the training data. That’s why the predictive methods necessitate more data than descriptive techniques. So, to be more efficient and get the best separation criterion aimed to predict the belonging class. the number of classes is initially known, if it equals to two, we are talking about a binary classification, otherwise it is a multi-class classification [17]. In this paper we focus on the application of supervised learning algorithms specially K-nearest neighbors (KNN), Support Vector Machine (SVM) and Random Forest Classifier (RFC) to identify the cell cycle and non-cell cycle regulated genes.

2.1 KNN:

KNN is a supervised learning algorithm, used for classification and regression analysis [18]. It requires a training dataset which is a set of labelled data. KNN assumes that similar objects are near to each other, so it considers the k learning samples whose entry is closest to the new entry x. KNN does not make any assumptions on the initial data. This algorithm doesn’t learn a model, it memorizes the training instances. To predict classes, KNN defined similarity by calculating the Euclidian or Hamming distance between the new data point and the labelled data point and identified the majority of nearest neighbors. The most popular and familiar distance measure is Euclidian distance, it is calculated using this formula:

$$d\,(p,q)= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \ldots + (qn + pn)^2}\ (1)$$

The accuracy depends on the optimal choice of k. To select the K that reduces the number of errors, we tested k in[2,3,4,5,6,7,8,9,10,11,12,13].

2.2 SVM:

SVM is a supervised learning algorithm defined for classification tasks, regression and outlier's detection[19]. It is very used in medical diagnosis to evaluate cancer, detect cardiac arrhythmia. SVM classifier is used to predict a binary qualitative variable by finding good decision boundaries. A decision boundary is a hyperplane in an N-dimensional space of N features that distinctly classifies the data points. The hyperplane separates the dataset into two different spaces belonging two different classes. The objective is to find the plane by mapping the data then trying to maximize the distance between the hyperplane and the two classes data points what we call maximize the interclass distance and minimize the intra-class distance. These hyperplanes defined by Vapnik and Cortes as a linear function with maximal margin between the vectors[20]. It is described by the following formula:

$$f(x)=w^{t\cdot}x+b(2)$$

The best separator is a hyperplane corresponding to f(x) = 0. It is valid if x is on the margin. The kernel function is noted as measure to remodel the training data and to map it into a higher dimensional space in order to make it linear. For many domains, the transformation of the features data into a space with a higher number of dimensions a makes easier the separation of the classes with a linear function [21]

- Polynomial Kernel:

This Kernel is generally used to map a hyperbolic surface to a plane, is non-stationary, that's why it is well adapted to normalized datasets. It is often used in image processing[22]. The polynomial degree d and the constant term r should be carefully chosen.

$$K\,(x_i, x_j) = (\sigma\, x_i{}^T x_j + r)\ d,\ \sigma > 0 \qquad (3)$$

The test of the polynomial kernel was conducted at r=0.0 and degree=3

- Radial Basis Functions Kernel:

RBF as a Gaussian function is a universal kernel. We use this kernel when we miss a prior knowledge about the data [23]. It is a flattened function which measures the distance using an exponential function

$$K\,(x_i, x_j) = (\sigma\ \| x_i x_j \|^2)\ \sigma > 0 \qquad (4)$$

To get the best accuracy we should choose the convenient gamma and the adequate c parameters . That's why we tested the following c and σ in:

c= [1, 10, 100, 1000]Andσ = [0.01, 0.001, 0.0001]

- Sigmoid Kernel:

It is also known as the Hyperbolic Tangent Kernel. This kernel should be a positive definite function to obey the Mercer's theory [24]. Sigmoid kernel is semi-definite for certain parameter values

$$K\,(x_i, x_j) = tanh\,(\sigma\, x_i{}^T x_j + r)(5)$$

The parameters that control the mapping are gamma is an input data scaling and r is a shift. These two parameters should be chosen appropriately to not warp the results.

2.3 RFC:

RFC is a hierarchy classifier, a meta estimator. It is a set of tree-shaped learning algorithm. The Random Forest Classifier fits an ensemble of decision trees classifiers on randomly chosen subgroup of the training data set [25]. It sums the votes from separate decision trees to determine the class of the test point. Here we varied the number of trees in the forest equal from 4 to 19, The number of features to consider when looking for the best split is very important. To find the best classifier we tested the auto, sqrt and log2.

3. DEEP LEARNING TECHNIQUE

Deep learning is a subset of machine learning. It isa class of models that are built with layers of neurons. Artificial Neural Network (ANN) inspired from the imitation of the human brain physiology, and the way it works. These models have been applied to many fields such as thecomputer vision, audio recognition, Bioinformatics and medical image diagnosis,have produced results comparable to human experts.

An artificial neural network is a network of artificial neurons. Artificial neurons are equivalent to biological human neurons. A natural neuron takes in a neural impulse electrical signal from the axons of other neuron and makes action. This approach of input output is a measurable function. This function does some computations to the input signal, then produce an output. Artificial neurons are also known as perceptron. The biological electrical signals correspond to numerical values. Deep neural networks are a multi-layer artificial neural network, where the layers number specifies the deep of the model[26].

The first layer of a neural network is called the input layer. It takes input values or signals and delivers them to the next layer. The total of the input layer's neurons should equalize the set of features number of the dataset. The mathematical representation of a biological neuron is the fundamental unit of a neural network. They analysis the entering data, then determine how to characterize it by proportionate a weight for each input [27].

The value of the coming will multiplied by the weight that is specifically accorded to it. First, these weights are randomly initialized, then the neural network discovers more about what type of input data leads to a particular result. Based on the errors that the network makes, it adjusts the weights. This step is called the model training. A higher weight will allocate to the important input parameter value [28].

The hidden layers are a set of layers placed between the input layer and the output layer. A hidden layer can provide different transformations to the data coming from his previous layer.

The last layer of a neural network is called an output layer. The output of all the previous hidden layer is given to this layer as input. The number of neurons of the output layer should be equal to the number of the desired classes.

Convolutional Layer is a particular layer modifying the input by a filter. In convolutional networks, numerous filters initiated to cut the data and map the parts one by one and learn different portions. A filter is a small-scale square matrix with random values. They are going be subjected to the data that will be illustrated as a matrix. Their job is to mining features and extract patterns in the training data. This matrix is going to be the output of this layer and will then be forwarded to the next layer [29].

Max Pooling is a method used with CNN. This task follows the convolution transformation. It decreased the resolution of a given output of a convolutional operation and reduces the number of feature parameters in the neural network.Max Pooling can also regularize and hold off the over fitting.

The algorithm of neural networks and how neural networks learn is back propagation. It consists on the update of the weights by considering the actual result and the desired output. So, it is the practice of adjustments of the weights based on the error rate that will give the most efficient decrease to the cost function [29]. By learning, it determines the phenomenon of understanding which weights and biases should be chosen to minimize the error measure.

4. PROPOSED METHOD

Our objective is to classify the cell cycle regulated gene using the DNA microarray images. To this end, we mainly have four parts as shown on fig.1. First, we start by collecting and understanding the data. Second, we prepare our dataset by deciding on the data that we are going to use for analysis and construct the required data. Third, we select modeling technique that we will use. We apply the SVM, KNN, RFC and Deep learning techniques to classify cell cycle-regulated genes and non-cell cycle regulated-genes. Finally, in the evaluation phase, wemeasure the models' performances for the CCRG classification [30]

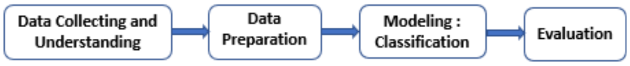


Figure 1. The steps of our proposed approach

In the first and second step, we use the database of the Yeast Cell Cycle project. This dataset is reported by Spellman et al. on the author's website (http://genome-www.stanford.edu/cellcycle/) [12]. To build the dataset, we are interested with the microarray's images of cdc15 temperature-sensitive mutant. The cell culture was shifted at 37 C and they tracked the evolution of genes for 3.5h so we have 24 microarray images for 24 different experience progression's moments. These images are RGB color sizing 1024*1024 (Figure.2).

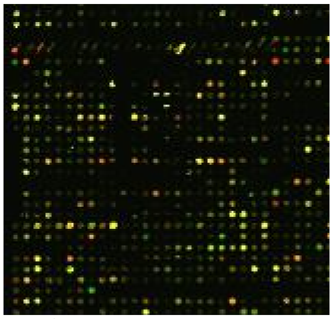


Figure 2. Microarray Image

Each image is related to a raw data enclosing a set of information: genes' name, ratio, gene's location, etc. To process these images and classify the genes, we follow the process shown in Figure 3.

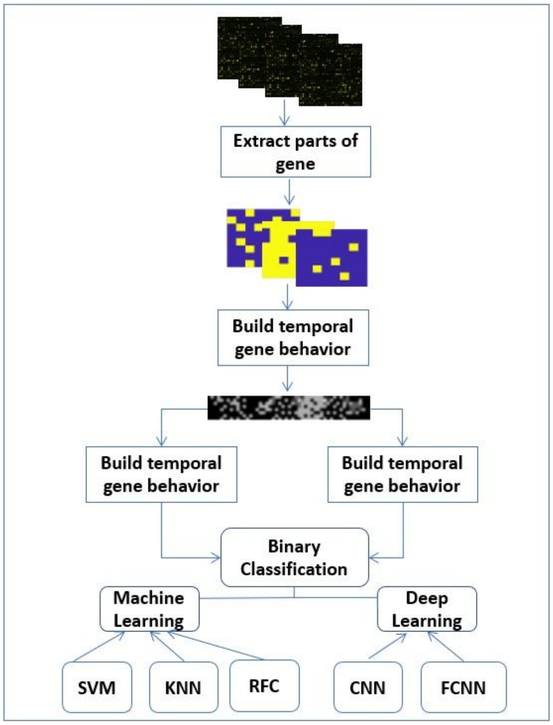


Figure 3.CCRG Classification process

To generate the gene's attitude, we extract the segments of each gene from the 24 microarrays images. Then we form the gene's progression from these parts by sequencing them, in order to visualize the gene evolution overall the 290 min. We pick up as a result, an image sizing (8*192) as shown on figure.4.

Figure4. Gene temporal progression image

To identify the CCR and NCCR genes, we exploit the new gene's images processed on the previous phase. Then we use the machine learning algorithms previously mentioned which are SVM, RFC and KNN by varying the parameters. We also propose two deep learning models a Convolutional Neural Networks (CNN) and a Connected Neural Networks(FCNN) to outperform the classification. These models take as input, the labeled image data representing the two classes cell cycle and non-cell cycle regulated genes.

The Fully Connected Neural Networks is built by the following layers: The input shape is 8*192 and the first hidden layer is a Flatten. The next layer was a dense, with 512 units, is applied with a hyperbolic tangent activation function. We follow this layer by a dropout one to decrease the over-fit on the neural networks. The next two hidden layers has the same construction followed by a dropout layer. The last hidden layer is a dense layer activated by a softmax function.Fig.5 shows the architecture of our FCNN model.[30]

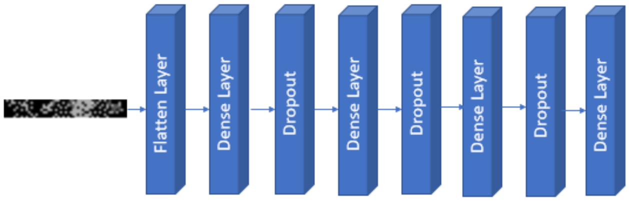


Figure 5. Model summary of the FCNN

The proposed Convolutional Neural Networks model is as follows: The input shape is 8*192 of pixels and up to 3 color channels. The first hidden layer is a convolutional layer with 16 filters activated by a Relu function. This layer is then followed by a Max Pooling layer with a pool region of size 2 by 2. Then a flatten layeris applied. The hidden layer has a dense Layer with 64 units, followed by another dense layer with 32 units with a hyperbolic tangent activation function. Finally, a dense layer with 2 units is activated by a Softmax function. Fig.6 shows the architecture of our CNN model.[30]

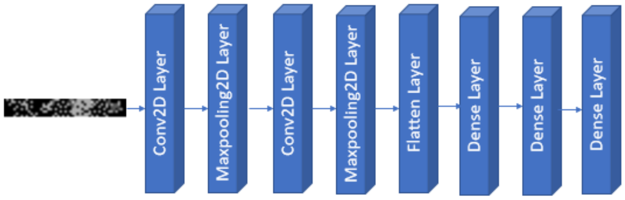


Figure 6. Model summary of the CNN

The FCNN model needs an enormous number of connections and network parameters. The peculiarity of CNN is that it considered that an image is combined of details, or features, which provides an abstract about the whole image.

5.RESULTS

In our work, we aim to classify cell cycle-regulated genes and non-cell cycle regulated-genes. For this purpose, we build a dataset composed of labeled images describing genes behavior. To classify genes, we take 284 images per class and split our data to train set and test set. To assure that our model has got randomly and independently the two sets, we use cross validation function to split the input data into 80% for training and 20% for testing. To Fit and compile the deep learning model, we train it for a fixed number of epochs. The number of samples per gradient is updated, the batch size is 8. A portion of the training data is used as validation data. The model will set apart 30% of the training data and will evaluate the loss and other metrics at the end of each epoch. To quantify the quality of predictions and to evaluate the results of our models and methods, we choose the accuracy and the Loss Function as scoring parameters and four indexes including the sensitivity or recall, precision, F1 score and support given by equations from 6 to 10. For the loss function, we choose the log loss, named also Binary Cross-Entropy Loss. It increases as the predicted probability diverges from the actual label. A perfect model might have a log loss of 0

$$Accuracy= \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Sensitivity = recall = \frac{TP}{TP+FN} (7)$$

$$Precision = \frac{TP}{TP+FP} (8)$$

$$F1 = 2* \frac{Precision*Sensitivity}{Precision+sensitivity} \quad (9)$$

$$LogLoss= -(y \log(p)+(1-y)\log(1-p)) (10)$$

Where TN, TP, FN and FP are short for true negative, true positive, false negative and false positive. y is a binary indicator of whether class label is the correct observation's classification and p is the model's predicted probability. To verify the convergence of our models, we plot the accuracy and loss graphs. The best K classifier parameter is K= 6. Table I shows the evaluation results. The method of the K Nearest Neighbors allow us to get an accuracy of 92.10 %.

Table 1: knn's Evaluation metrics

	Precision	Sensitivity	F1-score	Support
CCRG	0.86	0.97	0.93	59
NCCRG	0.96	0.87	0.91	55
Accuracy (%)	92.10			

Similarly, for the SVM classifier, we tested the four different functions kernel RBF, Sigmoid, Linear and Polynomial. The best accuracy is improved by polynomial kernel.The parameters of the best classifier are c=10, gamma=0.01(Table II)

Table 2 : Linear kernel evaluation indexes

	Precision	Sensitivity	F1-score	Support
CCRG	0.98	0.98	0.95	57
NCCRG	0.92	0.91	0.95	57
Accuracy (%)	94.73			

The method of the Random forest classifier gives us an accuracy of 88.31 % as shown on Table V. The best classifier is provided by those parameters'values:max_features=auto and n_estimators=16.

Table3 : RFC evaluation indexes

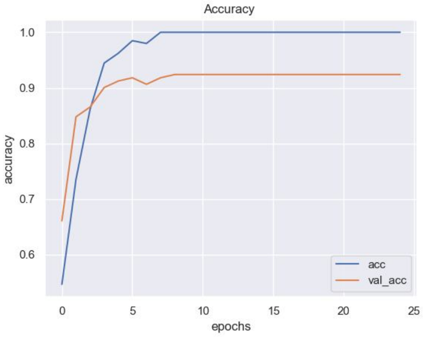
	Precision	Sensitivity	F1-score	Support
CCRG	0.83	0.47	0.60	40
NCCRG	0.89	0.98	0.93	174
Accuracy (%)	88.31			

To evaluate the robustness of our method we compared our classification results to Liu et AL's results. This comparison is illustrated in the Table VI and is made upon three criteria that are precision, recall, and F1.

Table 4 : Results comparison

	Our results			Liu et al 's results		
	Precision	recall	F1	Precision	recall	F1
KNN	0.92	0.92	0.92	0.7	0.8	0.55
SVM	0.95	0.95	0.95	0.8	0.9	0.5
RFC	0.88	0.88	0.87			

Based on the previous comparisons, we conclude that our approach is more consistent than previous studies. Best results were obtained for SVM-Polynomial with accuracy rate 94.73%. We find a better accuracy because, in our approach, we take into account the temporal continuity constraint and the behavioral progression of the cell regulated genes and non-cell cycle regulated genes.



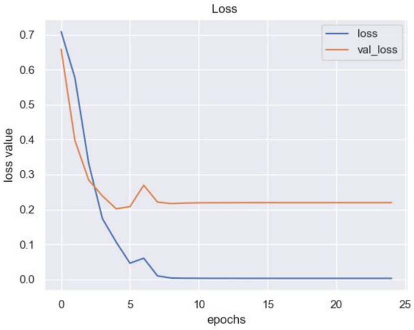


Figure 7. CNN Model Accuracy and Loss

Our CNNmodel, accomplish the classification accuracy of 92.39%. Then we tested the FCNN model on the same database.

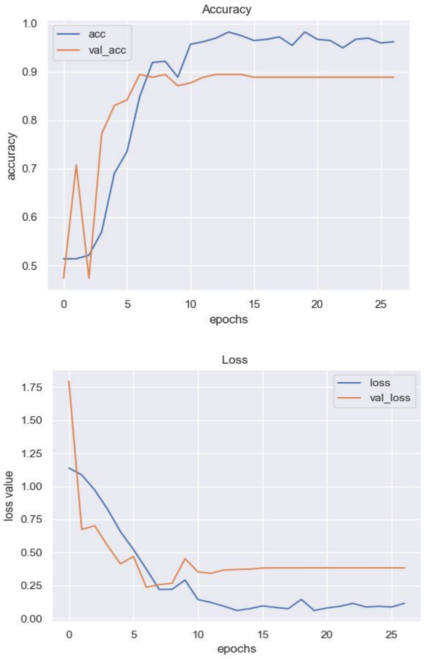


Figure8. FCNN Model Accuracy and Loss

Our proposed FCNN architecture reaches an accuracy of 90.64%. To evaluate the method's robustness, we compare our classification results to Liu et AL's results. This comparison is illustrated in the Table V and is made on the precision's criteria. The experiments show that our approach outperforms the other model by a margin of 22.39 per cent.

Table5 : Results comparison

	Accuracy	
Our results	CNN	92.39%
	FCNN	90.64%
Liu et al 's results DLGene	CNN: 70%.	

The powerful point of our approach consists on the classification of genes based only on images data, in which we modeled the gene behavior. For image recognition, the CNN model is more efficient than the FCNN model because it applies to input images a convolution action to well describing it.

6. CONCLUSION

In this paper, we apply a new approach to process the microarrays images and classify them using the machine learning techniques and deep learning methods to recognize cell cycle and non-cell regulated genes. The novelty resides particularly in building a temporal behavior gene images from the microarrays and using it in the classification to consider the temporal progression .For classification using Deep learning we proposed a CNN and FCNN models, we got an interesting accuracy. Our CNN model reaches an accuracy rate of 92.39%. The Comparison made with usual research shows that our classification models outperform other methods by a margin of 22.39%. For classical machine learning classification, we used KNN, RFC and SVM, we got very encouraging accuracies. Best results were obtained for SVM-Polynomial with accuracy rate 94.73%. Comparisons also show that our machine learning methods outperforms other methods by a margin of 14.73%.

REFERENCES

1. A.WC. Liew, H. Yan and M. Yang. **Pattern recognition techniques for the emerging field of bioinformatics**, Pattern Recognition Societ., vol. 38, pp. 2055–2073, 2005.
2. M.P.S. Brown, W.N. Grundy , D. Lin, N. Cristianini, C.W. Sugnet, T S. Furey, M. Ares, Jr. and D Haussler. **Knowledge-based analysis of microarray gene expression data by using support vectormachines**, PNAS. vol. 97, pp. 262–267, January 4, 2000.
3. A. Gupta, H. Wang and M. Ganapathiraju. **Learning structure in gene expression data using deep architectures, with an application to gene clustering,** in Washington D.C, November, 2015.
4. M. Molla, M. Waddell, D. Page, and J. Shavlik. **Using Machine Learning to Design And Interpret Gene-Expressed Microarrays,** AIMag, vol. 25, no. 1, p. 23, Mar. 2004.
5. D. Johansson, P. Lindgren and A. Berglund, **A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription,** Bioinformatics, vol. 19, no. 4, pp. 467-473, 2003.
6. A. Huttenhofer and J. VogelGavin. **Experimental approaches to identify non-coding RNAs,** Nucleic Acids Research, vol. 34, pp.635–646, 2006
7. L. MW1 and N. WS.**Machine learning applications in genetics and genomics ,** Nat Rev Genet, pp 321–332, Jun 2015
8. J. Bacardit and X. Llorà**. Large□scale data mining using genetics□based machine learning,** GECCO, vol. 09, pp. 740-741, July, 2009
9. **V. Uslan and Đ.Ö. Bucak, Microarray Image Segmentation Using Clustering Methodes,** Mathematical and Computational Applications, vol. 15, pp. 240-247, 2010.
10. L. MW1 and N. WS.**Machine learning applications in genetics and genomics ,** Nat Rev Genet, pp 321–332, 2015
11. B. Maresca, G. Kobayashi and H. Yamaguchi, **Molecular biology and its application to medical mycology.** .
12. P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R.Iyer, K.Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B.Futcher.**Comprehensive Identification of Cell Cycle–regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization**, Molecular Biology of the Cell, vol. 9, pp.3273–3297, December 1998.
13. C. Liu, P. Cui and T. Huang. **Identification of cell cycle-regulated genes by convolutional neural network,**Combinatorial Chemistry & High Throughput Screening, vol. 20, pp. 603-611, 2017
14. P. Mamoshina, A. Vieira, E. Putin and A. Zhavoronkov. **Applications of Deep Learning in Biomedicine**, Molecular Pharmaceutics, vol. 13, no. 5, pp. 1445-1454, 2016.
15. R. Touati, A.E.Oueslati, I. Messaoudi and Z. Lachiri.**SVM Helitrons recognition based on features extracted from the FCGS representation**,ATSIP'2017, vol. 9, pp.1–6, May 2017
16. A.WC. Liew, H. Yan and M. Yang. **Pattern recognition techniques for the emerging field of bioinformatics**, Pattern Recognition Societ., vol. 38, pp. 2055–2073, 2005.
17. P. Mamoshina, A. Vieira, E. Putin and A. Zhavoronkov. **Applications of Deep Learning in Biomedicine**, Molecular Pharmaceutics, vol. 13, no. 5, pp. 1445-1454, 2016.
18. M. Molla, M. Waddell, D. Page, and J. Shavlik.**Using Machine Learning to Design and Interpret Gene-Expression Microarrays,** AIMag, vol. 25, no. 1, p. 23, Mar. 2004.
19. J. Bacardit and X. Llorà. **Large□scale data mining using genetics□based machine learning**, GECCO, vol. 09, pp. 740-741, July, 2009.
20. A. Huttenhofer and J. VogelGavin.**Experimental approaches to identify non-coding RNAs,** Nucleic Acids Research, vol. 34, pp.635–646, 2006.
21. R. Touati, A.E.Oueslati, I. Messaoudi and Z. Lachiri.**A Novel methodology for Helitron's classification by SVM based on time frequency representation**, ICCAD, vol. 17, pp.281–286, Junuary 19-21 2017.
22. P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov. **Applications of deep learning in biomedicine**,Molecular Pharmaceutics, March 24,2016.

23. Y. LeCun, Y.Bengio, G.Hinton, **Deep learning**. Nature 521,pp. 436–444 doi:10.1038,nature14539 , 2015.
24. E. Orr and B. Martin. **Evolving A-type artificial neural networks,** Evolutionary Intelligence, vol. 5, no. 1, pp. 3-22, 2011.
25. D. Bychkov et al.**Deep learning based tissue analysis predicts outcome in colorectal cancer**, Scientific Reports, vol. 8, no. 1, 2018.
26. E. Putin, P. Mamoshina, A. Aliper, M. Korzinkin, A. Moskalev, A. Kolosov, A. Ostrovskiy, C. Cantor, J. Vijg and A. Zhavoronkov.**Deep biomarkers of human aging: Application of deep neural networks to biomarker development**, Aging, 8(5), pp.1021-1033, 2017
27. M. Oquab, L. Bottou, I. Laptev and J. Sivic. **Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks**, Computer Vision and Pattern Recognition, 2014.
28. D. Karimi, H. Dou, S. Warfield and A. Gholipour.**Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis**, Medical Image Analysis, vol. 65, p. 101759, 2020
29. H. Lahmer, A. E. Oueslati and Z. Lachiri.**DNA Microarray Analysis Using Machine Learning to** Recognize Cell Cycle Regulated Genes, 2019 International Conference on Control, Automation and Diagnosis (ICCAD), Grenoble, France, 2019, pp. 1-5, doi: 10.1109/ICCAD46983.2019.9037868.
30. H. Lahmer, A. E. Oueslati and Z. Lachiri, **Classification of DNA Microarrays Using Deep Learning to identify Cell Cycle Regulated Genes**,2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, 2020.