ISSN No:-2456-2165

Optimized Gene Classification using Support Vector Machine with Convolutional Neural Network for Cancer Detection from Gene Expression Microarray Data

Vishwas Victor; Dr. Ragini Shukla
Department of IT & CS., Dr. C. V. Raman University Kota, Bilaspur, India

Abstract:- There are numerous approaches for handling microarray gene expression data since new feature selection techniques are constantly being developed. To create a new subset of pertinent features, feature selection (FS) is utilized to pinpoint the essential feature subset. The model that used the informative subset projected that a classification model generated solely using this subset would have higher predicted accuracy than a model developed using the whole collection of attributes.

We offer an analytical approach for cancer classification and developed a model using Support Vector Machine as classifier and after that Convolutional Neural Network in the aspect of Deep Learning. The outcome received in the context of the proposed model is very impressive and accurate.

Keywords:- *Feature selection; Optimization;Classification; Support Vector Machine (SVM); Deep Learning; Machine Learning; Convolutional Neural Network (CNN).*

I. INTRODUCTION

This A microarray expression experiment records the expression levels of thousands of genes simultaneously; each gene is a segment of DNA that carries all the information needed to make several kinds of proteins in our body. The main methods used in these experiments are explained by authors that either multiple monitoring of each gene under various conditions, or evaluating each gene in a single environment but in different types of tissues, particularly cancerous tissues [9]. One method for reducing classifier calculation errors is feature selection (FS), which removes noisy, redundant, and unrelated qualities from the original data set and selects important attributes. According to the authors, generally feature selection techniques fall into three main categories: wrapper, filter, and embedded models [7].

The microarray dataset technique faces two primary issues: an excessive number of genes relative to a lesser number of samples. The process of identifying relevant features from the data and displaying the higher dimension dataset with a condensed search space is known as feature selection (FS). The proper FS resolution for microarray data, however, is very difficult to determine because the sample size is smaller than the total number of genes. There are a number of things to take into account when reducing the dataset's dimensionality, [13]. Microarray gene expression studies frequently produce a large number of characteristics for a limited number of patients, resulting in a high dimensional dataset with a small number of samples. Gene expression data is extremely complicated and problematic; genes are connected with one other either directly or indirectly, making the classification process extremely tough and challenging. Typically, this means employing a precise and potent feature selection technique is required [10].

The field of Deep Learning (DL) is concerned with using deep networks for information processing. Convolutional Neural Networks (CNNs) are a type of artificial neural network that can extract local features from data. CNN assigns weights based on a single feature mapping, simplifying the network model and allowing for a reduction in overall weights. DL is designed to handle data using both supervised and unsupervised methods, with learning taking place on several layers of features and descriptions [17]. The processes of feature extraction and selection result in a distilled set of the essential characteristics that define the core characteristics of the data and categorize it. You can carry out this classification with or without supervision. While supervised learning uses data with output class markers, unsupervised learning uses information without output class labels. An algorithm illustrates the relationship between patterns in the input attribute variables and the associated descriptors in the output for supervised approaches [5]

The main objective of this literature to get better results with highest accuracy for feature selection of gene expression microarray data. Lots of works happened till date, but at this end for the humanity it is required to re-examine the datasets with better prediction related to identify the cancer form the infected gene expression microarray data. Here, we have proposed an embedded approach for Dimension Reduction and Classification using Support Vector Machine with Convolutional Neural Network (DRC-SVM-CNN). This approach has produced the result with better accuracy and prediction.

This literature contains the following sections: In Section 2, the Microarray Technology and Datasets described. In Section 3, Related works compiled by the researchers and their published articles are evaluated. In Section 4 Proposed Methods and model which are used to address the challenge of features selection, explained. In

Section 5, implementation and analysis explained in this section. Section 6 consists result analysis and discussion, obtained through proposed model. In Section 7, conclusion of the literature explained.

II. MICROARRAY TECHNOLOGY AND DATASETS

It is intended to apply a tissue or DNA sample to the chip's minuscule areas. Deoxyribonucleic acid, or DNA, is a type of genetic material found in nearly all living things, including humans. A throughput technique used in cancer research for illness detection and prognosis is microarray data. Microarray technology is the most straightforward method for identifying patterns in both normal and aberrant tissue. A silicon device called a microarray displays thousands of small dots. First, several genetic samples from unknown individuals—normal and abnormal—must be obtained. These are cautiously placed into the silicon chip's little openings. Following that, DNAs hybridize. Annealing primer is used to separate the RNA, mRNA, and tRNA from these DNAs, which are linked to their complementary DNA. A tagged cDNA is produced following the transcriptase enzyme's labeling of mRNA. After the tagged cDNA hybridizes, a green or red laser light is seen on the chip. Eventually, the intensity is calculated and microarray data is produced.

In the given figure 1, the process to create gene expression microarray data is illustrated. First, we collect the sample from the infected person, then by applying them through microarray silicon chip, we got the microarray data.

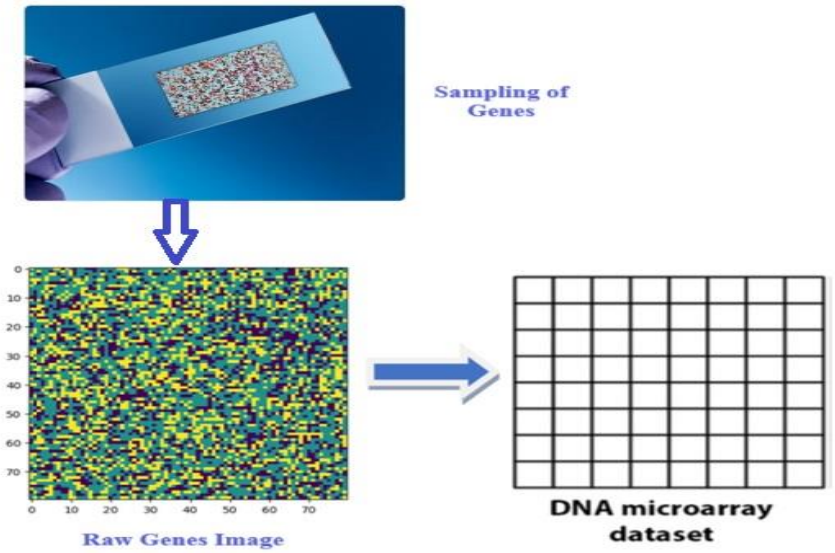

Fig. 1: Gene expression microarray data creation process

These are the address of gene expression microarray datasets repository from where we can get them for our experiments-

- https://www.ncbi.nlm.nih.gov
- http://csse.szu.edu.cn/staff/zhuzx/Datasets.html
- http://www.biolab.si/supp/bi-cancer/projections/
- https://web.stanford.edu/~hastie/CASI_files/DATA/leukemia.html
- https://jundongl.github.io/scikit-feature/datasets.html,
- https://vizgen.com/data-release-program/
- https://dataverse.harvard.edu/

Our analysis has been done on these datasets given in the Table I.

Table 1: Description of Microarray Datasets

Dataset	Features	Sample	No. of Classes
Breast	24,481	78	2
Prostate	12,600	102	2
Brain	12,625	21	2
CNS	7129	60	2
Colon	2000	62	2
DLBCL	4026	47	2
GLI	22,283	85	2
Ovarian	15,154	253	2
Leukemia	7070	72	2

III. RELATED WORKS

For choosing a set of microarray data, a minimum collection of features or genes (or subsets of features), a variety of strategies have been put forth by various scholars. After conducting a thorough literature review, we will have a brief discussion regarding current methods as they pertain to our suggested work.

Authors have proposed an algorithm based on Deep Neural Network using Convolutional Neural Network[1]. They implemented the technique to recover overfitting problem in DNN. Authors explained the complexity of the gene expression microarray datasets, as these datasets consists huge number of features. They described about their proposed improved-Deep Neural Network algorithm to find out the relevant features to predict the best solutions.

Authors have described the nature of gene expression microarray datasets, which is contains lots of information under it [2]. Gene expression microarray datasets are high dimensional datasets and it is challenging to get the most relevant features predict the probabilities of disease. Authors have proposed Deep Gene Selection (DGS) algorithm for the best performance to achieved classification of cancer. They compared various gene selection methods with DGS and got the best results from DGS with better accuracy and computation cost.

Utilizing sophisticated classification and prediction methods, microarray datasets, which contain information about the various gene expression patterns have been studied, in order to help with the crucial and difficult task of early diagnosis of chronic diseases like cancer. To address the high-dimensional microarray datasets problem and ultimately improve the accuracy of cancer classification, a hybrid filter-genetic feature selection strategy has been presented by the authors [3]. To further optimize and enrich the chosen features and increase the suggested method's capacity for cancer classification, a genetic algorithm has been used. According to the experimental findings, the suggested hybrid filter-genetic feature selection strategy outperformed a number of widely used machine learning techniques to obtain accuracy, recall, F-measure, and precision.

Gene expression profiling, or microarrays, evaluate and identify patterns and levels of gene expression in a variety of cell types and tissues. Author explained the Cancer classification using DNA microarray technology, which allows for the simultaneous intense treatment of hundreds of gene expressions on a single chip. The Latent Feature Selection Technique is proposed in the suggested method by the author, to shorten classification times and improve accuracy after deep learning algorithms are trained on microarray data to extract features [4]. This study used bone marrow PC gene expression data to develop the Artificial Bee Colony (ABC) feature selection method. Author found the tumor classification, performed using Convolutional Neural Networks has a better accuracy.

The authors have proposed a model using statistical methods and machine learning for feature selection. Support Vector Machine (SVM) was the one method for selection of the features subset [5]. Finding this genetic material from the entire genome is essential, and depending on how these few genes express themselves, a patient's condition can be classified as either having or not. Author highlighted the various challenges associated with the gene expression data in the presence of noise and missing values.

Authors have explained a deep feedforward approach for categorizing provided microarray cancer data into a set of classifications for eventual diagnosis [6]. Authors employed a 7-layer deep neural network design with different parameters on all datasets. The difficulties of very less sample size and high dimensionality have been solved by using a popular dimension reduction method named principal component analysis. The Min-Max methodology is used to scale the feature values, and the suggested method is verified on eight common microarray cancer datasets. A comparison with state-of-the-art approaches is conducted, and the effectiveness of the suggested methodology outperforms several of the existing methods.

Authors have explained some of the most recent feature selection approaches for analyzing the microarray datasets to produce better results [7]. Author described that the large number of features and limited quantity of samples, microarray data classification presents a significant task for machine learning researchers. Author offered an experimental assessment on the most significant datasets utilizing recognized feature selection methods, with the goal of facilitating comparative investigation by the research society rather than providing the optimal feature selection approach.

Authors have defined that comparison of the period of evaluation, classification accuracy, and possibilities to recognize the disease, as well as calculating the strictness of the positions of disease, the experimental outputs show that deep neural network classification execution can be better than present classification techniques [8]. The authors have explained about the necessary to take a close look at this topic, as well as the research findings and related issues, in order to achieve a perspective of the ailment categories. The author have described, Neural networks are the foundation of machine learning, which includes deep learning. Knowledge, or the ability to distinguish between different types of information and organize it into a manner that is easily comprehended by humans, has been retrieved from the massive amount of raw data. Cancer prediction is deep learning's primary function.

Authors have proposed a method on the basis of Support Vector Machine (SVM) and Mutual Information (MI). The authors explained that SVM is a technique for supervised learning that can handle challenging classification issues, whereas the informative genes can be determined using the mutual information (MI) among the genes and the class label. By selecting the most useful gene subset, the suggested method improves classification

accuracy while reducing the dimension of the input characteristics when compared to alternative methods [9].

Authors have developed A hybrid cancer classification method by using number of machine learning techniques, including the easy-to-understand Decision Tree classifier that doesn't need a parameter, Grid Search CV (cross-validation) to optimize the maximum depth hyperparameter, and Pearson's correlation coefficient as a correlation-based feature selection and reducer. To assess the approach, seven common microarray cancer datasets are employed. The results show that the suggested technique chooses the most informative characteristics, boosts classification accuracy, and significantly reduces the number of genes needed for classification [9].

Authors have introduced an efficient feature selection technique that explores the nonlinear mapping skills of double RBF-kernels in conjunction with weighted analysis for obtaining feature genes from gene expression data [11]. Outperformed earlier approaches with comparatively higher accuracy, true positive rate, false positive rate, and less runtime, proposed method tested on four benchmark datasets with either two-class phenotypes or multiclass phenotypes. The authors explained, with a large number of genes and limited samples, the main difficulty in analyzing gene expression data is separating disease-related information from a vast amount of noise and redundant data. To solve this issue, removing unnecessary and duplicate genes by gene selection has been a crucial step.

Authors have suggested a method to significantly reduce the data source's dimension in order to increase the accuracy of decreasing dimensionality. The study classifies feature selection algorithms into three categories: semi-supervised, supervised, and un-supervised learning. The performance of multiple described methods in the literature was analyzed, as well as the recent efforts to reduce the features for tumor diagnosis [12].

Limited instances with a significant amount of imbalance between the classes can occasionally be found in gene expression data. This may restrict a classification model's exposure to examples from other categories, which may have an impact on the model's performance, explained by the authors. Authors have developed an algorithm for feature selection by employing numerous iterations of 5-fold cross-validation on various microarray datasets. The algorithms' performance varied according to the data and feature reduction method applied [14]. Authors suggested a method to identify cancer by using data preprocessing methods such feature selection, oversampling, and classification models. The six datasets that were evaluated in the study were oversampled using SVMSMOTE.

The authors have explained one of the best tools for addressing the identification of cancer, prediction, and treatment is microarray gene-based expression profiling techniques. In healthcare, microarray research is used for cancer research as well as illness prognosis and treatment [16]. The authors have described, Neural networks serve as the foundation for deep learning, which is a kind of automated learning. The data has been extracted from the extensive body of knowledge that is raw data, allowing for discrimination and easy comprehension within a framework.

Authors have developed a model named hybrid mSVM-RFE-iRF for gene selection and classification using Support Vector Machine (SVM), Recursive Feature Elimination (RFE) and improved Random Forest (iRF) and compared with Convolutional Neural Network (CNN). Deep learning algorithm applications are receiving a lot of interest as a means of resolving various AI-related problems. According to author, when compared to hybrid mSVM-RFE-iRF, the majority of experimental results on cancer datasets showed that CNN is more accurate and minimizes gene when it comes to cancer classification [17].

Our goal is to provide an effective model that can choose the smallest feature set with the highest degree of accuracy. To that end, we have taken into consideration the deep learning technique known as Convolutional Neural Network, executed on optimized genes collected by Support Vector Machine and Random Forest classifiers.

IV. PROPOSED TECHNIQUES

Finding the w and b parameters that define the hyperplane in a way that optimizes the margin which is the distance between the hyperplane and the closest data points from each class is the primary goal. Support vectors are defined as the set of these nearest neighbors. Followings are the steps taken by us for analysis of the gene expression data:

A. Dataset partitioning:

Three subsets of the dataset were created: a first set of features for training, a second set for validation, and a third set for testing. While the training dataset (first one) is used to train the classifier that is utilized (algorithm), the validation dataset is used to evaluate the performance of the classifier and is applied in the optimization efficiency. In order to assess the effectiveness of the entire feature selection and classifier procedure, the test dataset (final dataset) is employed.

B. Decreasing dimensionality of datasets:

By using measurement approaches to find features that were essential in the initial dataset for categorizing a target, the dataset is reduced.

C. Assigning values for specifying parameters:

The feature selection approach creates meaningful feature subsets by selecting important characteristics from the original dataset after removing unnecessary, redundant, and noisy features. The potential feature subsets are evaluated to find out classification accuracy. An explanation of the study technique is provided in Figure 2:

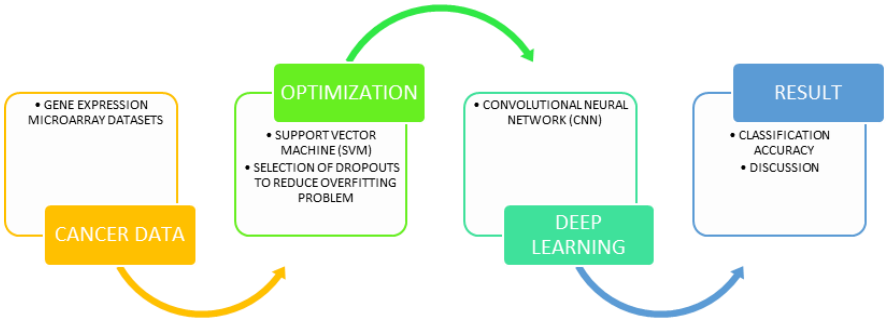

Fig. 2: Proposed (DRC-SVM-CNN) algorithm flow diagram for classification

The conclusion will consequently be minimum feature size and highest classification accuracy in order to acquire the best feature subsets (standard deviation and average number of features).

D. Applied Methodologies

Here we are explaining the applied methodologies for the analysis of the research.

➢ *Support Vector Machine*

The foundation of Support Vector Machines (SVMs) is a mathematical formulation that looks for the best hyperplane to divide two data classes. An SVM's principal mathematical expression is as follows:

Given a dataset with input data points x_i and their corresponding labels $y_i \in \{-1,1\}$

where i =1, 2, …, N, an SVM seeks to find a hyperplane defined by:

$$w \cdot x + b = 0 \quad (1)$$

- w is the weight vector that is orthogonal to the hyperplane.
- x represents the input features of a data point.
- b is the bias term that shifts the hyperplane away from the origin.

Finding the values of w and b that maximize the margin, the distance between the hyperplane and the closest data points from each class, is the aim of defining the hyperplane. Support vectors are those closest data points.

The following is a formulation for the SVM optimization:

$$Y_i (w \cdot x_i + b) \geq M, \text{ for } i = 1, 2, \ldots, N \quad (2)$$

Where:

- M is the margin (The distance apart the two parallel hyperplanes are).
- w and b are the parameters to be optimized.

➢ *Convolutional Neural Network*

Convolutional neural networks, or CNNs, are a kind of deep learning models that are mainly utilized for tasks involving images, however they can also be used with other kinds of data. Figure 4 illustrating the typical architecture of Convolutional Neural Network (CNN).

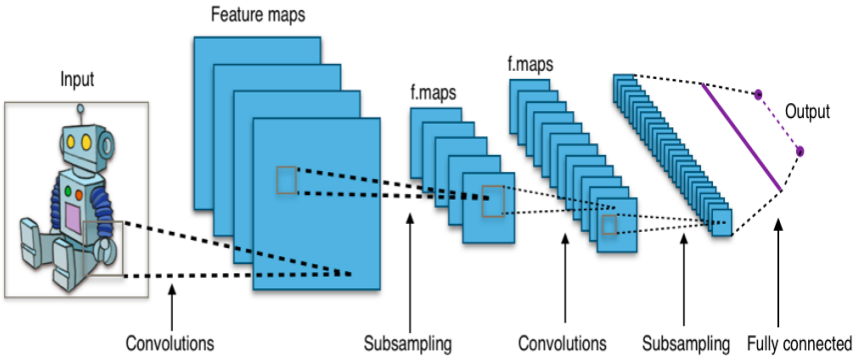

Fig. 3: Convolutional Neural Network Architecture

The following are the essential elements of a typical CNN's mathematical expression:

➢ *Convolution Operation*

Convolution, the primary function of a CNN, is the process of extracting local characteristics from input data by swiping a filter, sometimes referred to as a kernel, across the data.

The mathematical representation of the convolution operation for a 2D input, such as an image, is as follows:

$$(I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) \cdot K(m, n) \quad (3)$$

Where:

- I input.
- K is the convolutional kernel.
- (i,j) represents the pixel location in the output feature map.
- (m,n) represents the pixel location in the kernel.

➢ *Activation Function*

After convolution, an activation function (commonly ReLU - Rectified Linear Unit) is applied element-wise to introduce non-linearity into the model. Mathematically, this can be represented as:

$$ReLU(x) = \max(0, x) \quad (4)$$

➢ *Pooling Operation*

Pooling (e.g., max-pooling or average-pooling) is used to down-sample the feature maps, reducing the spatial dimensions and the number of parameters.

Mathematically, max-pooling can be represented as:

$$\text{Max Pooling}(x) = \max(\text{neighbors of } x) \quad (5)$$

➢ *Fully Connected Layers*

After several convolutional and pooling layers, CNNs often have one or more fully connected layers to make high-level predictions.

A fully connected layer can be expressed mathematically as a matrix multiplication followed by an activation function, typically a softmax for classification tasks.

➢ *Loss Function*

The error separating the genuine values from the predicted values is measured by the loss function. Depending on the task at hand, a loss function such as cross-entropy for classification or mean squared error (MSE) for regression should be used.

➢ *Back propagation and Optimization*

CNNs are trained by varying the model's parameters (weights and biases) to minimize the loss function through the use of optimization algorithms such as stochastic gradient descent and backpropagation.

➢ *Proposed Algorithm (DRC-SVM-CNN)*

- Collect Gene Expression Microarray Datasets
- Preprocess the datasets for analysis
- Apply Support Vector Machine for classification
- For optimization, select dropouts to reduce the overfitting problems
- Apply and tarin the autoencoder to reduce feature dimension
- Again, train SVM classifier on the reduced features and analyse
- Create a Convolutional neural network model for taking reduced features as input Utilizing the encoded features, train the convolutional neural network to perform classification without overfitting.

V. IMPLEMENTATION AND ANALYSIS

To get experimental results, we have implemented our proposed method using Python 3.9 version and Jupyter Notebook. We have used computer system with Intel i5, 1.80 GHz processor, 16GB RAM and Windows 11 environment. The outcomes achieved by the experimental results are provided here as we have collected.

Let us consider the following:

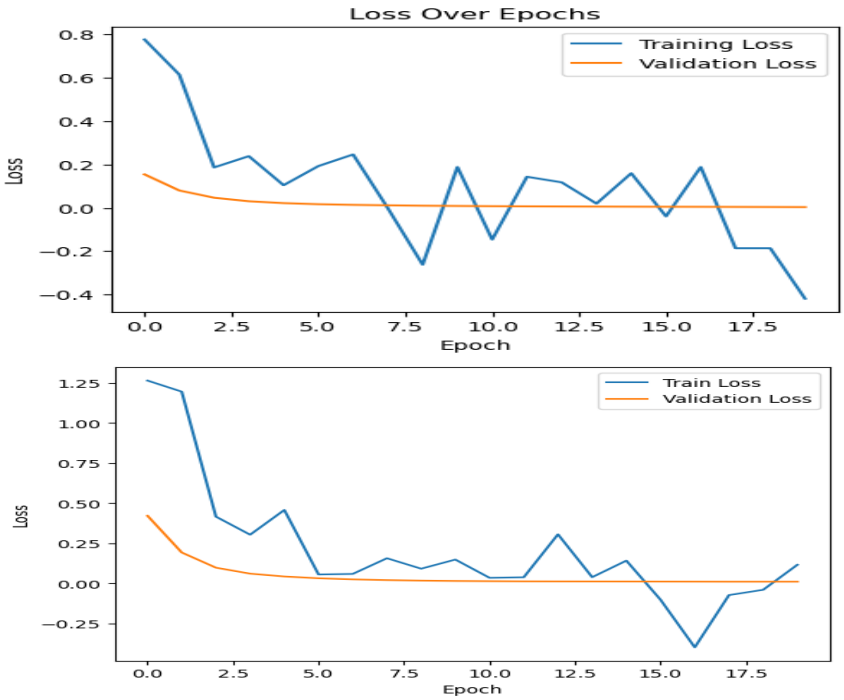

Fig. 4: Observing the Test and Validation loss over the Epochs

ISSN No:-2456-2165

Figure 4 represents the training and testing loss evaluation, where we have found that validation loss is approximately at 0 constantly, whereas training loss is fluctuating between 0.2 and -0.2 in each epochs due to the high dimensional data cause overfitting. Here we can say that the training loss is also very less and negligible, so that the training of model is in appropriate condition. Without implementation of dropout selection and SVM, it found very uncertain.

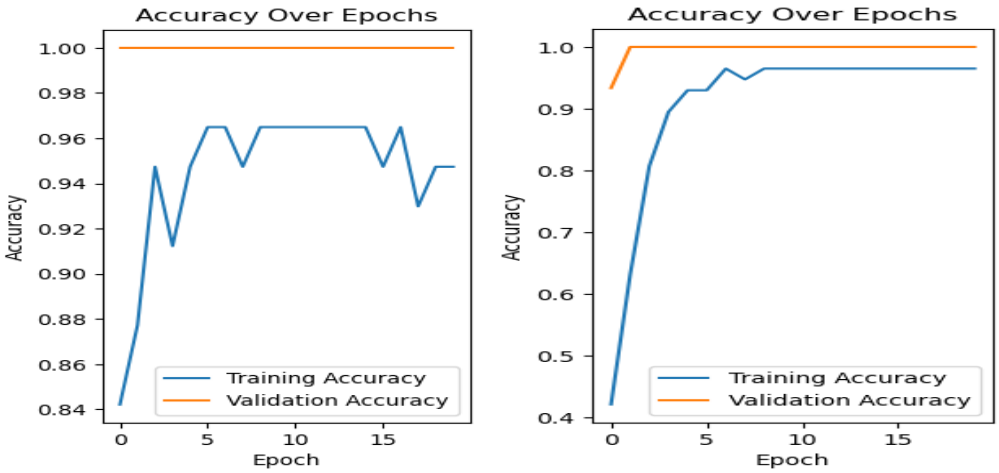

Fig. 5: Result Accuracy Representation of Data

Figure 5 represents the accuracy over all epochs and we observed that the validation accuracy is 100% and training accuracy is at nearly 97%. So, that in this graphical representation our proposed model has get fulfilled the expectations. In this regards we have analysed that the proposed model is outperformed, one of the better solution for optimization and classification of gene expression microarray data.

VI. RESULTS ANALYSIS AND DISCUSSION

According to the proposed model the classification accuracy of original feature sets are very impressive. Tensorflow, Numpy, Scikit-Learn are used in our proposed program. To achieve these results, datasets were divided for experimental purpose into two parts that are Training and Testing, such as 80:20 ratios. Sensitivity, Specificity, and Accuracy are checked for performance evaluation.

A. Sensitivity

Sensitivity measures the proportion of actual positive cases that the model correctly identifies as positive. It is calculated as:

Sensitivity = (True Positives) / (True Positives + False Negatives)

Sensitivity is important when you want to minimize false negatives, as in medical diagnoses, where missing a true positive (e.g., a disease) could be critical.

B. Specificity

Specificity measures the proportion of actual negative cases that the model correctly identifies as negative. It is calculated as:

Specificity = (True Negatives) / (True Negatives + False Positives)

Specificity is important when you want to minimize false positives.

C. Accuracy

Accuracy is a measure of overall correctness. It calculates the proportion of all cases (both positive and negative) that the model classifies correctly. It is calculated as:

Accuracy = (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives)

Accuracy provides a broad view of a model's performance but may not be the best metric when class imbalances exist or when certain types of errors are more costly than others.

D. Results from Selected Datasets

The following results have described the experimental findings for each microarray dataset. After optimization of gene expression microarray data by selecting dropouts and applying Support Vector Machine with Convolutional Neural Network, our model gets trained and analysed by calculating three tests for Sensitivity, Specificity, and Accuracy. Here we can see the collected results with classification accuracy almost above 99% on average. So, that the performance of the proposed model is very much suitable for the prediction of the disease identification through gene expression microarray data. The performance assessments of the complete microarray dataset are mentioned in Table 2.

Table 2: Datasets used in the experimental study and obtained results

Datasets	Genes Selected	Sensitivity	Specificity	Accuracy
Breast	4,896	97.68	97.86	**99.79**
Prostate	2,520	97.12	98.35	**98.55**
Brain	2,525	98.03	98.41	**99.03**
CNS	1425	97.62	98.89	**99.15**
Colon	400	98.93	98.3	**100**
DLBCL	805	97.62	97.79	**98.64**
GLI	4,456	96.64	98.73	**99.62**
Ovarian	3,031	97.68	98.2	**99.74**
Leukaemia	1414	98.68	98.81	**99.91**

VII. CONCLUSION

In recent years, there has been an increase in the occurrence of challenging diseases, the significance of producing microarray data from different tissues, and the analysis of this data. As we know the challenge of uncertainty of the microarray gene expression data due to huge number of features and very small number of instances. To overcome this problem, it is required to optimize the data and reduce the irrelevant and redundant data from datasets, having huge number of features or dimension. We have worked in this research, to optimize the gene expression microarray data using our proposed method by implementation of SVM and dropout selection, thereafter obtained optimized data has processed through convolutional neural network. This proposed model is named by us as DRC-SVM-CNN can be identified as embedded approach for dimension reduction and classification using support vector machine (SVM) with convolutional neural network (CNN). The SVM classifier provides enhanced results after creating a new feature set. This result has been contrasted with classifier accuracy beyond sum dataset. Selected genes from nine cancer datasets were to be chosen using a convolutional neural network (CNN). Prior to classifying and choosing the right genes in each dataset, it is imperative to pick informative genes accurately. Using our model, we have obtained on average 99% accuracy among all nine datasets. Certainly, we can say that, this model (DRC-SVM-CNN) has produced all time better results in the area of gene expression microarray datasets for cancer classification.

REFERENCES

[1]. O. Ahmed and A. Brifcani, "Gene Expression Classification Based on Deep Learning," 4th Scientific International Conference Najaf, SICN, pp. 145–149, 2019.

[2]. R. Alanni, J. Hou, H. Azzawi, and Y. Xiang. "Deep Gene Selection Method to Select Genes from Microarray Datasets for Cancer Classification," BMC Bioinformatics, vol. 20(1), 2019.

[3]. W. Ali, and F. Saeed, "Hybrid Filter and Genetic Algorithm-Based Feature Selection for Improving Cancer Classification in High-Dimensional Microarray Data," Processes, 11(2), 2023.

[4]. H. Z. Almarzouki, "Deep-Learning-Based Cancer Profiles Classification Using Gene Expression Data Profile," Journal of Healthcare Engineering, 2022.

[5]. S. Arora, and S. Gupta, "Feature Selection of Gene Expression Data Using Machine Learning and Statistical Modelling," International Journal of Recent Scientific Research, Vol. 12, Issue, 10 (B), pp. 43334-43340, 2021.

[6]. H. S. Basavegowda, and G. Dagnew, "Deep Learning Approach for Microarray Cancer Data Classification," CAAI Transactions on Intelligence Technology, vol. *5*(1), pp. 22–33, 2020.

[7]. V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, J. M. Benítez, & F. Herrera, "A Review of Microarray Datasets and Applied Feature Selection Methods," Information Sciences, vol. ***282***, pp. 111–135. 2014.

[8]. V. Chandrasekar, V. Sureshkumar, T. S. Kumar, and S. Shanmugapriya, "Disease Prediction Based on Micro Array Classification using Deep Learning Techniques," Microprocessors and Microsystems, *77*, 2020.

[9]. C. Devi Arockia Vanitha, D. Devaraj, and M. Venkatesulu, "Gene Expression Data Classification using Support Vector Machine and Mutual Information-Based Gene Selection," Procedia Computer Science, vol. *47*(C), pp. 13–21. 2014.

[10]. H. Fathi, H. Alsalman, A. Gumaei, I. I. M. Manhrawy, A. G. Hussien, and P. El-Kafrawy, "An Efficient Cancer Classification Model using Microarray and High-Dimensional Data," Computational Intelligence and Neuroscience, 2021.

[11]. S. Liu, C. Xu, Y. Zhang, J. Liu, B. Yu, X. Liu, and M. Dehmer, "Feature Selection of Gene Expression Data for Cancer Classification using Double RBF-Kernels," BMC Bioinformatics, 19(1), 2018.

[12]. N. Mahendran, P. M. Durai Raj Vincent, K. Srinivasan, and C. Y. Chang, "Machine Learning Based Computational Gene Selection Models: A Survey, Performance Evaluation, Open Issues, and Future Research Directions," Frontiers in Genetics, Vol. 11, 2020.

[13]. Malibari, R. M. Alshehri, F. N. Al-Wesabi, N. Negm, M. Al Duhayyim, A. M. Hilal, I. Yaseen, and A. Motwakel, "Deep Learning Enabled Microarray Gene Expression Classification for Data Science Applications," Computers, Materials and Continua, vol. *73*(2), pp. 4277–4290, 2022.

[14]. O. O. Petinrin, F. Saeed, N. Salim, M. Toseef, Z. Liu, and I. O. Muyide, "Dimension Reduction and Classifier-Based Feature Selection for Oversampled

Gene Expression Data and Cancer Classification," Processes, vol. 11(7), 2023.
[15]. S. H. Shah, M. J. Iqbal, I. Ahmad, S. Khan, and J. J. P. C. Rodrigues, "Optimized Gene Selection and Classification of Cancer from Microarray Gene Expression Data using Deep Learning," Neural Computing and Applications, 2020.
[16]. Shyamala Gowri, S. Bhuvaneshwari, A. Abirami, R. Dilli Rani, K. N. Anirudh, and S. Keerthi Shree, "DNA Microarray for Cancer Classification Using Deep Learning," European Chemical Bulletin, pp. 1631–1641, 2023.
[17]. Q. Zeebaree, H. Haron, and A. M. Abdulazeez, "Gene Selection and Classification of Microarray Data Using Convolutional Neural Network," International Conference on Advanced Science and Engineering (ICOASE), October 2018.