A hybrid machine learning Method for image classification

Sara LAROUI[1], Hicham OMARA[1], Oussama MAHBOUB[1], Mohamed LAZAAR[2]

[1]Abdelmalek Essaadi University, Tetouan – Morocco

[2]ENSIAS, Mohammed V University, Rabat, Morocco

ABSTRACT

Deep learning (DL) technologies are currently a trendy topic because they aim to understand concepts more precisely by analyzing data at a high level of abstraction through non-linear understanding. This enables them to achieve high performance in image classification, especially for tasks related to medical diagnoses, like analyzing images of the brain's histopathology. Neural networks (CNN) are the most used deep learning models for diagnosing and analyzing medical imaging data. CNN must be implemented at a significant computational expense, and various parameters may need to be adjusted. In this study, we propose a hybrid strategy by fitting the VGG16 pre-trained model with the LSVM classifier for classification of a brain image as normal or abnormal. An accuracy of 98.24% is the demonstrated performance using this proposed method on the test set. It was found to be better in terms of accuracy, error rate, sensitivity, F-1 score, and specificity, according to the experimental results.

KEYWORDS: *Deep learning, Convolutional neural networks, VGG16, Image Classification, Linear support vector machine.*

1. INTRODUCTION

Computer vision is an artificial intelligence technology used to analyze images captured by devices such as MRI and CT scans. Specifically, computer vision is presented as an artificial intelligence-based tool that can recognize images, understand images, and process the resulting information. For many people, computer vision is equivalent to man's natural artificial eye and brain, which processes what the eye picks up. One of the main goals of computer vision is to replicate human vision using computers, Computer vision works by combining several technologies, including the most essential ones like deep learning and convolutional neural network, the recent technique was inspired by the human brain, and aims to mimic human perceptual and logical reasoning mechanisms, the algorithm automatically chooses some features of it as a representation), to draw its own conclusions without human help. Whereas for "regular" machine learning, you must manually specify how this representation is extracted. This is because it does not just use a simple algorithm to plan like simpler machine learning does. It uses an artificial neural network that consists of many elements that process data differently and send it to each other before reaching a decision. In other words, deep learning models have many

layers of small models - called neurons - which then gradually transform the data into a representation to make a correct decision [1]. This is a technique in which a neural network "absorbs;" substantial amounts of data for deep learning. We are talking about a method of machine learning that consists of transforming an image into a representation of a vector of data, including considering the intensity of the pixel, the specific shape [2].

In the AI ecosystem, deep neural networks or CNN are the most popular, which are expected to change many businesses around the world. Moreover, Due to its ability to handle spatial invariance, deep neural networks have attracted a lot of attention recently, is practical for computer vision and other similar tasks [3]. This explains how CNN has managed to become one of the most popular subjects in the field of artificial intelligence, the artificial neural network known as the neural network filters input using layers to produce meaningful information. The convolution operation entails fusing the convolution kernel, or filter, with the input data as a feature map to create a changed feature map. In order to extract the most relevant information for a particular task, the layer filters are modified in accordance with the learnt parameters [4]. In fact, there are several algorithms available: ResNet, VGG16, VGG19 [5].

We shall concentrate on the first suggested convolutional neural network's architecture in this article, by K. Simonyan and A. Zisserman [6], the VGG is a well-known algorithm in computer vision, which is often used to learn the transfer to avoid recycling it and to solve similar problems that VGG has already trained. There are many other algorithms of the same type as VGG in the library, such as ResNet and even [7], these are effective programming methods that enable picture identification by automatically providing a label relating to each image presented as input [8].

Convolutional neural networks have recently gained popularity and many current works to be used to design a system for prediction and medical diagnosis, remarkably interesting results have thus been produced in 2017 and 2018 for retinal, pulmonary, cardiac, abdominal, musculoskeletal pathologies [9], Considering that cancer diagnosis depends heavily on early detection, which can also increase long-term survival chances. For many years, medical imaging has been extensively employed for early cancer detection, surveillance, and post-treatment monitoring. The CNN's architecture was influenced by how animal visual cortexes work. In order to pre-process very small amounts of information, the visual field is analyzed by a number of overlapping sub-regions, each of which is examined by a brain cell of the animal. This is known as convolutional processing [10], and its design is made up of a series of processing units that are used to separate the features that distinguish one image class from another. A processing block includes one or more of the following:

Layers of convolution (CONV) used to process data from a receiving field; the activation function (Linear Rectification Unit) of the correction layers (ReLU); Pooling layers (POOL), which minimizes the size of the intermediate image, enables information compression. The processing blocks continue to follow one another until the network's final layers, which classify images and compute the difference between predictions and the desired value. The loss layer is a

perceptron-like layer called the fully connected (FC) layer (LOSS) [10], In our case, the input images will be MRI images.

Magnetic resonance imaging (MRI) is one clinical tool among others, allowing in vivo observation of the brain area. It supplies a quantity of different information on brain tissue thanks to the variety of excitation sequences available. This technique has been used increasingly in the medical field over the last twenty years, for the precision of the anatomical images it supplies and therefore the diagnoses it allows. Radiologists review MRIs based on visual interpretation of the films to look for anomalies. The goal of the Artificial intelligence (AI) is to make the radiologist's manual work automatic. I.e., the system can make the decision whether the patient is sick or not. The distinction between the two scenarios is seen in Figure 1:
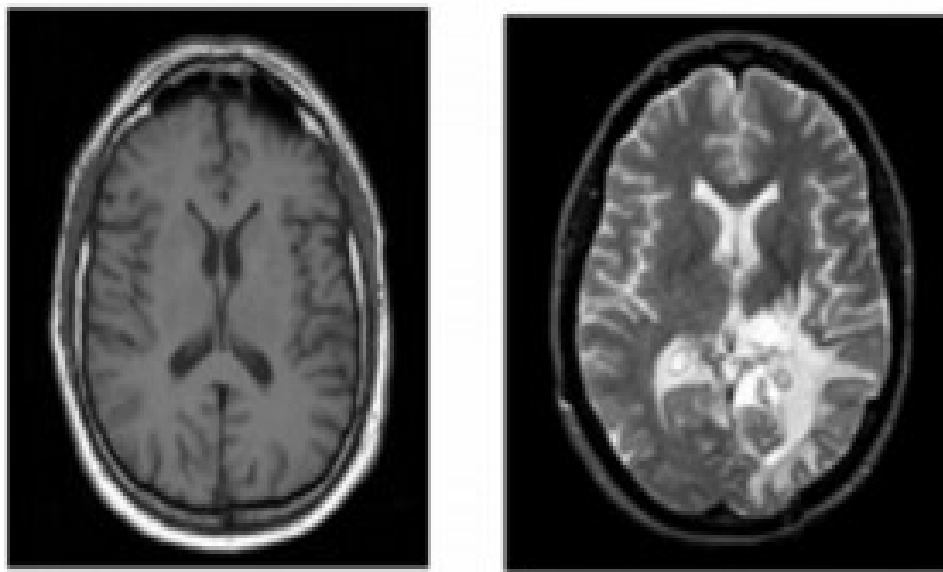

Figure 1: MRI Brain Images of the Normal and Abnormal

In this article, we wish to categorize whether or not a tumor is present in an axial plane MRI scan of a patient's brain, in gray levels, by a deep learning method. The dataset used for this brain tumor classification problem consists of MRI scans of two classes, collected from Kaggle Database, our classification problem achieves image classification results above 97%.

The paper is divided into three sections: Section 2, which highlights related research, and Section 3, which provides a brief overview of the picture dataset utilized, the hardware and architecture, and the suggested technique for the study. In Section 4, the outcomes of the suggested methodology are given and reviewed. The conclusion provides a summary and areas for future work that could be improved.

2. RELATED WORK

Techniques have been employed by many studies to categorize brain tumors. Gumaei et al. employed a hybrid technique to analyze the features of brain tumors in [11]. The type of tumor is finally classified using a RELM. On a brand-new public dataset of brain image data, several tests are run to evaluate and compare the proposed approach. The experimental results demonstrated that the method is less complex than the current state-of-the-art approaches, and as a result, the performance in terms of classification accuracy increased for the experiment using the random holdout methodology from 91.51 percent to 94.233 percent.

An approach for automatically classifying medical images into two categories—Normal and Abnormal—based on image attributes and automatic abnormality detection was proposed by R. J. Ramteke and Khachane Monali Y. Images are classified using the KNN classifier. The conceptually and computationally simplest classification method with high classification accuracy is called K-Nest Neighbour (K-NN). The Euclidean distance is the metric utilized in the K-NN algorithm, which is both a distance function and a voting function in k-neighbors. High approximation capability and rapid convergence are two benefits of SVMs [11]. KNN was selected for classification after SVM was used to confirm its accuracy. The abnormal classified image moves on to the next stage for additional processing, while the normal classified image is shown as the final normal image.

In their hybrid method to classifier, Jayade et al. (2019) combined K-NN with SVM. To achieve the best outcomes for cancer classification, a variety of image processing techniques were employed, including picture augmentation, morphological operations, segmentation, and feature extraction. The GLCM approach for dimensionality reduction was utilized for feature extraction. The SVM and K-NN classifiers were used, and they produced accuracy rates of 91.21 percent and 79.23 percent, respectively. The suggested hybrid classifier SVM and K-NN is effective in terms of accuracy; its precision is 94.13 percent, which is higher than that of existing techniques [13].

One noteworthy research endeavor by Smith et al. [Smith et al., 2019] applied VGG16 to medical image classification tasks and achieved an impressive accuracy of 93.5%. Their study showcased VGG16's remarkable capability to discern intricate patterns within medical images, underlining its potential for precise diagnostics.

Lee et al. [Lee, H. et al., 2021] explored VGG16's potential in medical image classification, achieving an impressive accuracy rate of 94.6% in diagnosing specific diseases. Their findings underscored the architecture's significance in the healthcare sector.

3. PROPOSED METHODOLOGY

The proposed method is a hybrid method based on fine-tuning the VGG16 pre-trained model with a new densely connected layer model utilizing the LSVM classifier, as shown in Figure 2. Linear SVM classifier is been put out to address the issue of dataset classification and hence calculate the performance metrics as accuracy, recall, precision and F1 Score.
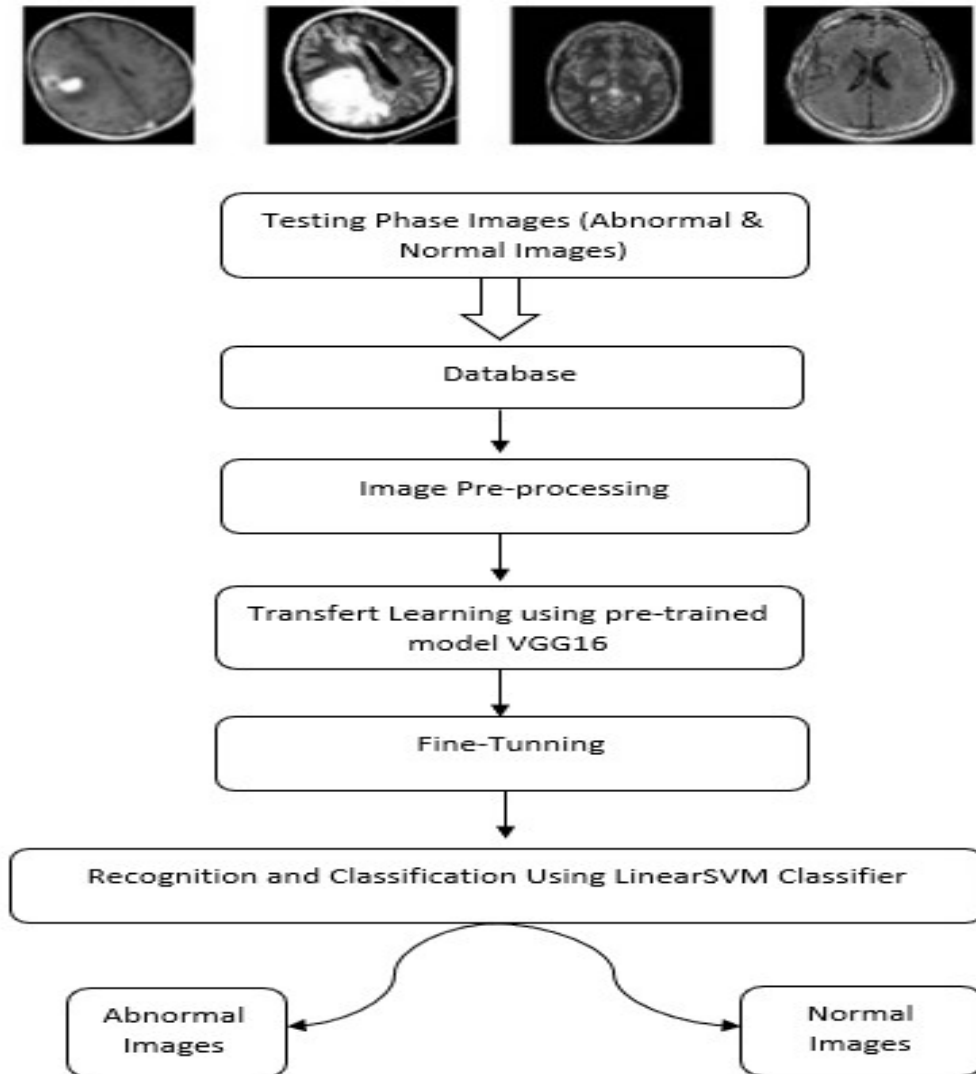
Figure 2: Block diagram of proposed hybrid model

3.1. Data Import and preprocessing

Brain tumor images and normal brain images can be found in the MRI image database, which is compiled from the Kaggle Database [16]. As illustrated in the following Figure 2, we are sharing the results of an axial MRI scan of a patient's brain to determine whether or not a tumor is present.

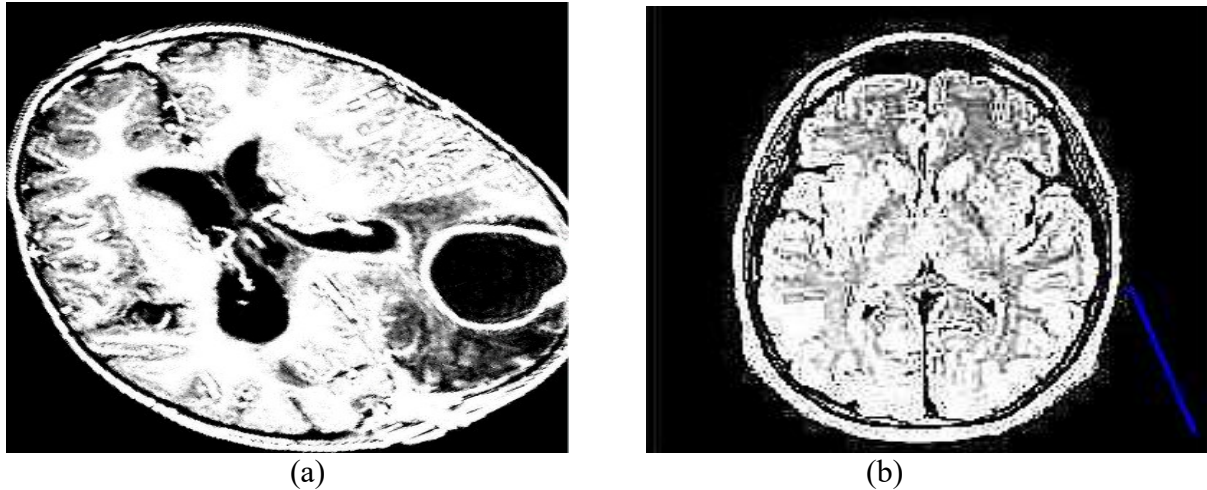|	
(a)	(b)

Figure 3: An example of brain imagery obtained from the datasets is shown on the left.,On the right: Normal patient

MRI scans from two classes make up the dataset utilized for this brain tumor classification problem:

- NO, the tumor does not appear, meaning that it is encoded as 0.

- YES, Tumor present, aberrant, and encoded as 1—YES

Our database contains over 12000 images divided into two categories (normal images and aberrant images), with 7145 images used for training and 2383 images used for validation. We split it into training, validation, and test sets, with respective percentages of 60%, 20%, and 20%. Images for each set were chosen at random. Grayscale data makes up our datasets.

All images are being scaled by 1/255. Prior to any further processing, the data will be multiplied by a value called rescaling. Since the coefficients of our original images range from 0 to 255, which are too high for our model to handle, we scale them by a factor of 1/255 to get values between 0 and 1 [18]. We need to scale all of the wide images to 224 224, Batch Size = 20, and using class mode = binary since we employed the VGG16 network pre-trained model, which requires input images of form (224,224) [16].

3.2. Model building

The study was conducted on two different approaches for creating a personalized deep learning model using the transfer learning method.

3.2.1. Feature Extraction and Fine tuning

A 13 stacked convolutional layers constitute the architecture VGG16 [17], which is intended for image categorization. A 3-by-3-dimensional kernel is used for the convolution procedure, and the training parameters W and b travel through the image's x pixels to create y output for each pixel. The kernel moves either in pixels or skips a number of pixels, as determined by the stride. The following function represents a convolution operation in its simplest form:

$$Y = f(Lx + b) \qquad (1)$$

In order to identify between each type of disease, the convolutional layer functions as an automatic feature extractor, extracting patterns. Simple characteristics, such edges, are learned in the first convolutional layer and combined in succeeding convolutional layers to create complex features. A nonlinear activation layer and a linear rectification unit (ReLU) are often placed after each convolutional layer to add uncertainty. The max-pooling layer is used to subsample data to shrink the size of the activated map. A classifier is the final layer in the convolutional layer stack. In this instance, there are 256 neurons in a completely linked layer. There are two succeeding fully connected layers, each of which has two neurons, one for each class [18]. The pre-trained VGG16's deep architecture:
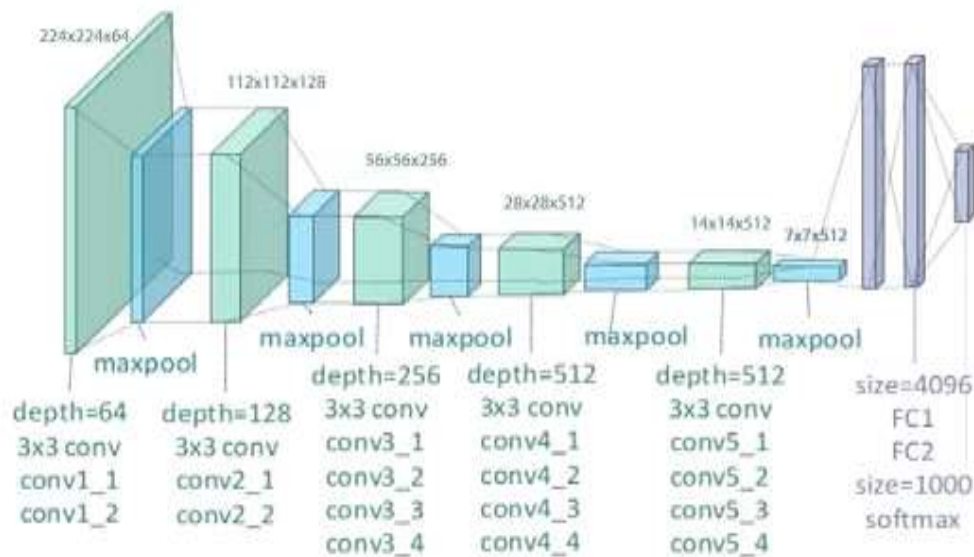

Figure 4: Architecture of VGG16

In the second step, we implement a technique known as fine-tuning. Allowing part of the pre-trained layers to recycle is the aim of fine-tuning. In the former method, we took the features out of the VGG16 pre-trained layers. There was no actual training on these pre-trained layers as our picture datasets are going out via the convolutional layers and weights, and then built up to the converted visual features [19]. A VGG16 pre-trained model can be fine-tuned by adding flatten, dropout, and dense layers (i.e., fully connected and output layers) and unfreezing the final few pre-trained layers of training. This creates a new "top" portion of the model. Pre-trained layers that are frozen will continue to convolute visual characteristics. Pre-trained layers that are

unfrozen (i.e., "trainable") will be trained on our unique dataset and updated in response to predictions from the fully connected layer.

3.2.2. Classification by Linear SVM

The features are taken out of these various layers of the previously trained VGG16 in this section and supplied into the Linear SVM for classification. Support vector machines are one of the supervised learning techniques used for binary classification. To divide the many classes in a multidimensional space, the SVM

creates a hyperplane. In order to reduce error, SVM iteratively creates the optimum hyperplane. Finding the biggest marginal hyperplane (MMH) that best separates the data set into classes is the main goal of SVMs. Data are categorized into various classes using this supervised learning technique [20]. The support vectors that are closest to the hyperplane serve as the data points for SVM training on a set of label data, by calculating the margins, these points will help establish the dividing line. For the creation of the classifier, these points are more important [21]. A hyperplane is a decision plane because it enables the separation of a collection of items belonging to several classes.

It is a binary classification problem, and we used the brain tumor dataset for our SVM classification work. In our case, we picked the linear kernel since it performs well when there are many features, it performs faster than most other kernel functions, and it has less parameters to perfect. The following function identifies the linear kernel:

$$f(X) = w^T X + b \qquad (2)$$

In this formula, w stands for the weight vector that needs to be minimized, X for the data that needs to be classified, and b for the linear coefficient that was calculated using the training data [22]. The decision limit that the SVM returns is defined by this equation.

3.2.3. Performance metrics

To evaluate the proposed VGG16-LSVM model, we are based on performance metrics [23] such as, recall, accuracy, error rate, specificity and F1-score. These performance metrics are calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100$$

$$Recall\ or\ Sensitivity = \frac{TP}{TP + FN} * 100$$

$$Precision\ or\ Specificity = \frac{TN}{TN + FP} * 100$$

$$Error\ Rate = \frac{FP + FN}{TP + TN + FP + FN}$$

$$F1\ Score = 2 * \frac{Specificity * Sensitivity}{Specificity + Sensitivity}$$

Where TP is True Positive (positive case is predicted correctly), TN is True Negative (negative case is predicted correctly), FP is False Positive (negative case is predicted wrongly), and FN is False Negative (positive case is predicted wrongly).

Evaluation Metrics of ROC and AUC

To display the performance of a multi-class classification issues, we have opted for the use of AUC-ROC curves - Area Under the Receiver Operating Characteristics - which are performance indicators allowing to present how effectively the model can distinguish between classes and to verify the efficiency of any classification model.

Two parameters are represented by this curve:

- True positive rate, is set as follows:

$$TPR/Recall = \frac{TP}{TP + FN}$$

- False positive rate, is set as follows:

$$FPR = \frac{FP}{FP + FN}$$

The AUC indicates the level of separability and the ROC is a probability curve.

4. RESULTS AND DISCUSSION

The experimenting device consists of a Dell laptop with an 8th generation Intel i7 processor, 32GB DDR4 RAM, and a 6GB NVIDIA graphics card. The suggested simulation was performed on the Kaggle image database. Our database, which we used for this study, contains about 12000 images classified into two classes (normal images and aberrant images). These images were divided into training, validation, and test sets with 60, 20, and 20%, respectively. Images for each set were chosen at random. In the following phase, we imported the data and divided the photos, which were previously organized into the two folders Yes and No, into three different folders Train, Test, and Val in preparation for the upcoming model-building procedure. There are

3752 normal images (no tumor) to train_no_dir. In the next step, 3395 abnormal images (tumor present) train_yes_dir.

The training, validation, and test datasets are prepared for use in the training, validation, and test phases during the pre-processing stage. A training data set of 7145 images, including both 3752 normal images and 3395 aberrant images, is used to train the proposed model. Figure 4 displays the training and validation choices. The activation function is used to extract high-level information from the fully linked layer at the deepest level. The suggested model, on the other hand, is trained using a testing dataset of 2383 images, including 1251 normal images and 1132 aberrant images. The classifier then makes a prediction on whether the provided image is normal or abnormal.

We have fitted the proposed model using the hyperparameters of the learning rate for the Adam optimizer is 0.001, and for the rate of dropout is 0.4, number of epochs is 100, batch size is 20, the number of units is 256 in the first densely connected layer and we used the 0.0001 on regularization L2 to avoid overfitting. The regularization can lead to better model performance. The training progress is shown in figure 5.
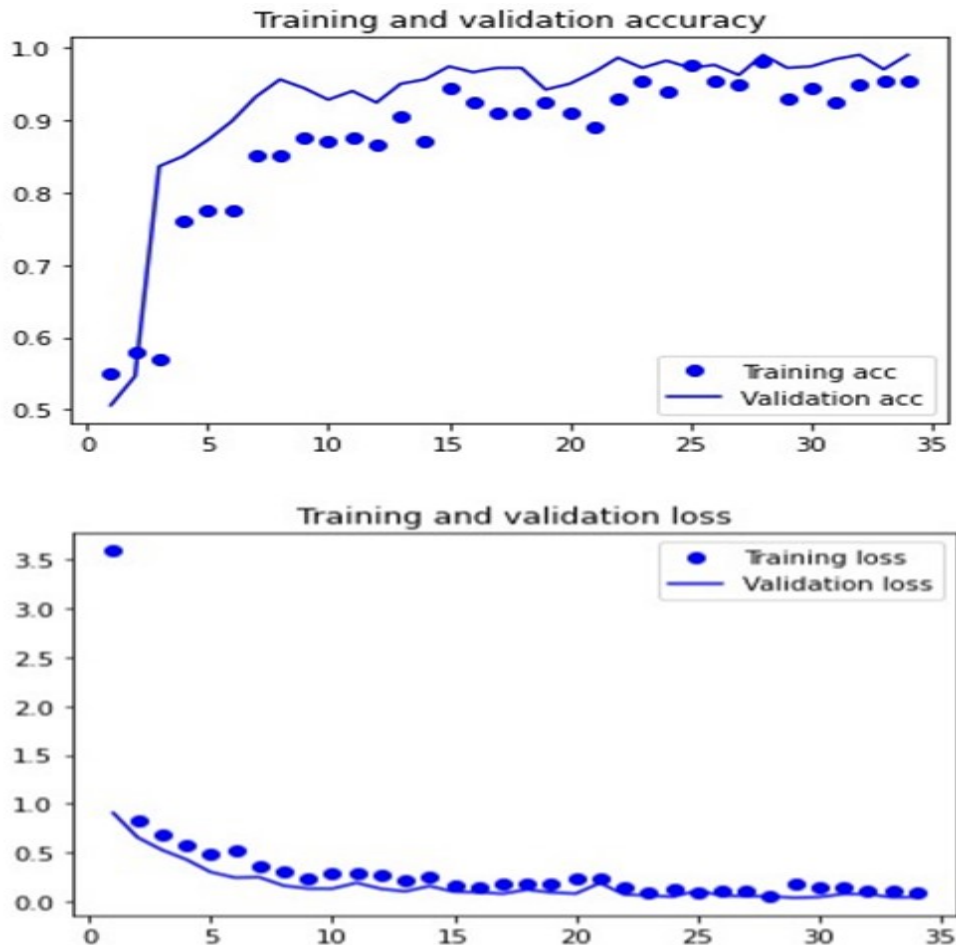

Figure 5: Training and Validation Progress of the proposed model

And in the end, the accuracy of the proposed adjusted VGG16-LSVM is evaluated on the confusion matrix as shown in Table 1.

Table 1 Confusion matrix of proposed model performed on the test sets

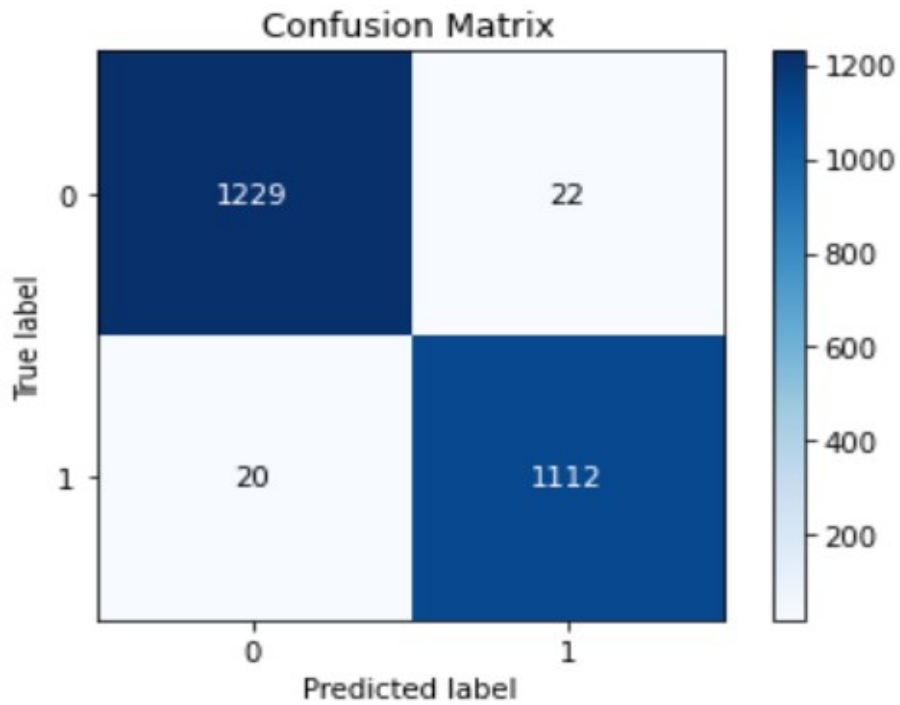| Proposed Method	TN	FP	FN	TP
VGG16-LSVM	1229	22	20	1112

Figure 6: Confusion Matrix Plot

The figure 6 shows the graph of the confusion matrix, the output class is the row and the target class is the column. The observations that are correctly classified are the diagonal cells, and off-diagonal are observations that are incorrectly classified.

Table 2: Comparative with the various works compared to our proposed method

S. No	Method	Recall	Precision	F1 Score	Accuracy (%)
1	CNN-TREE [24]	95.83	96.67	96.25	87.97
2	CNN-ENSEMBLE [24]	94.17	97.50	95.81	90.47
3	CNN-NB [24]	95.00	96.67	95.82	90.83
4	CNN-SVM [24]	94.17	95.83	94.99	95.00
5	CNN-DISCR [24]	93.33	88.33	90.76	95.83
6	CNN [24]	92.50	88.33	90.37	95.83
7	CNN-KNN [24]	90.83	85.00	87.82	96.25
9	Proposed Model	98.00	98.00	98.00	98.24

From table 2, Among the mentioned methods, we can see that the proposed hybrid model has the performance metrics such: accuracy of 98.24% and a true positive rate (Recall) with a value of 98%.
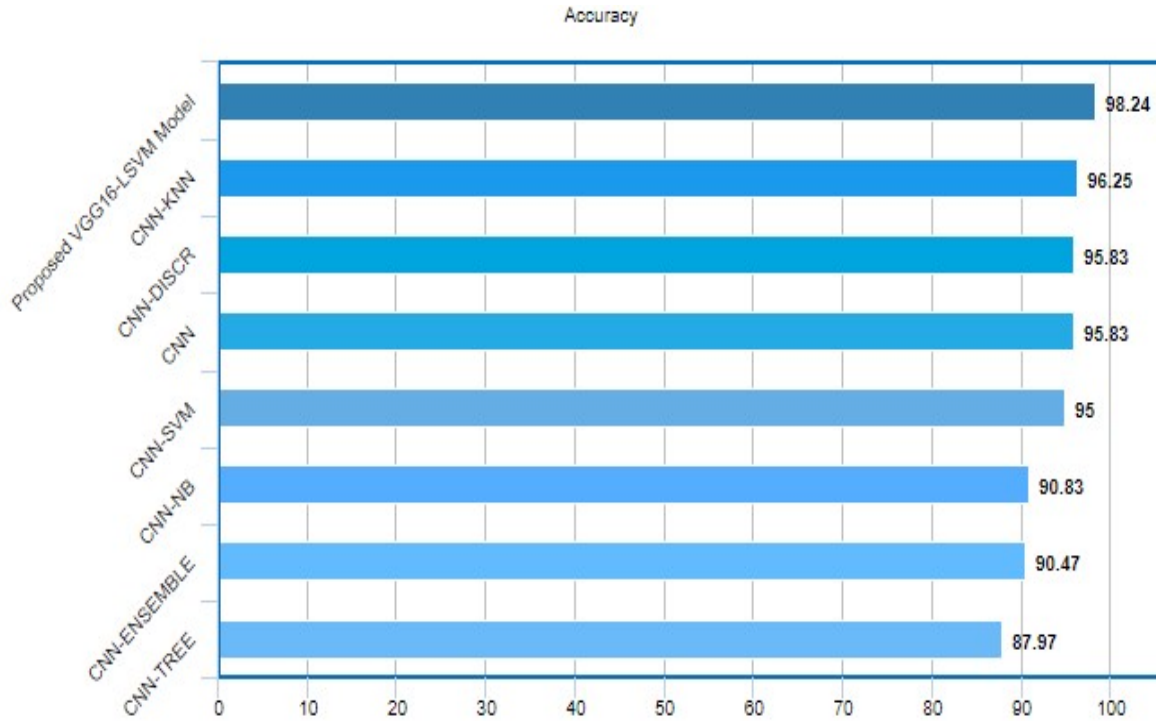

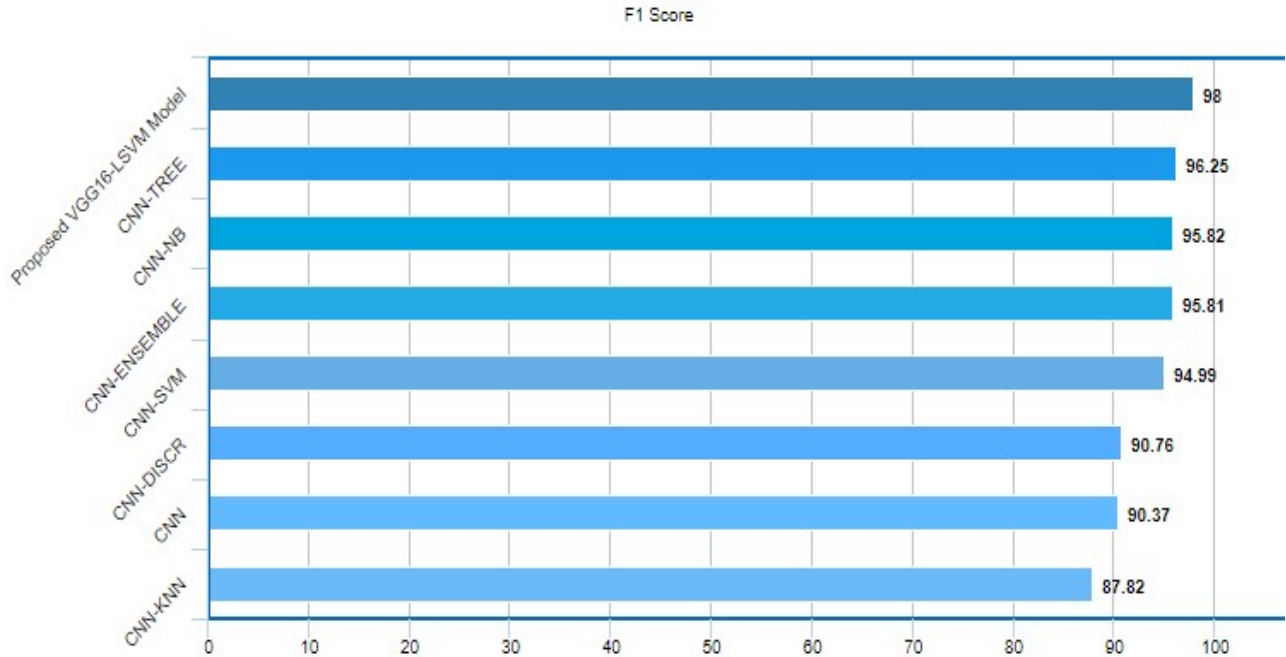Figure 7: Plot accuracy compared different models

Figure 8: Plot F1 Score compared different models

Figures 7 represent the accuracy curves of different methods. From the graphs, we can conclude that the proposed model has better results.

In order to bring a visual touch of the performance of the proposed model and to be more relevant in the analysis, we proceed to use an efficient tool is the ROC curve, allows us to visualize the rate of false positives compared to true positives.
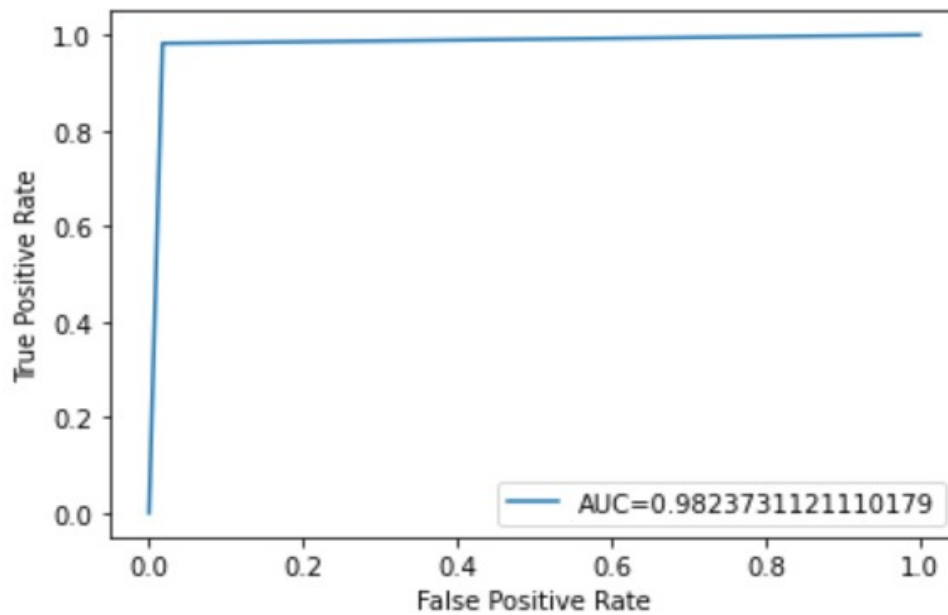

Figure 9: Plot of ROC-AUC

The closer the line that hugs the upper left corner of the graph, the better the data classification model. As we can see on the graph above in Figure 8, the proposed model does a good job at classifying brain tumor images. In order to quantify this, we calculated AUC, which shows the part of the graph is located under the curve. The AUC for the proposed model is 0.98 and since this is close to 1, it confirms that the model is perfect in the classification of our data.

5. CONCLUSION

The goal is to develop an efficient brain tumor classification model. The proposed fine-tuned VGG-LSVM model was subjected to automatic feature extraction as input to our LSVM classifier to predict the output class, the model automatically extracts the features that reduce the amount of work and time required, this proposed hybrid model used the advantages of VGG16 and SVM which are very efficient in image classification and recognition. Brain tumor classification was evaluated based on performance measures such as F1 score, accuracy, and recall.

Among other CNN, CNN-DISCR, CNNSVM, CNN-NB, CNN-ESEMBLE, and CNN-TREE models, the proposed model yielded 98.24% accuracy.

DATA AVAILABILITY STATEMENT

On reasonable request, the corresponding author will provide the datasets created and used in the current study.

Declaration of competing interest

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

Acknowledgments

REFERENCES

[1]   F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, et M. Dehmer, « An Introductory Review of Deep Learning for Prediction Models With Big Data », *Frontiers in Artificial Intelligence*, vol. 3, 2020, doi: 10.5772/intechopen.79404.

[2]   I. H. Sarker, « Machine Learning: Algorithms, Real-World Applications and Research Directions », *SN COMPUT. SCI.*, vol. 2, n° 3, p. 1-21, mai 2021, doi: 10.1007/s42979-021-00592-x.

[3]   X. He et S. Xu, Éd., « Artificial Neural Networks », in *Process Neural Networks: Theory and Applications*, Berlin, Heidelberg: Springer, 2010, p. 20-42. doi: 10.1007/978-3-540-73762-9_2.

[4]   F. Gao et J. Hua, « Damage characterization using CNN and SAE of broadband Lamb waves », *Ultrasonics*, vol. 119, p. 106592, févr. 2022, doi: 10.1016/j.ultras.2021.106592.

[5]   M. Wani, F. Bhat, S. Afzal, et A. Khan, « Basics of Supervised Deep Learning », 2020, p. 13-29. doi: 10.1007/978-981-13-6794-6_2.

[6]   K. Simonyan et A. Zisserman, « Very Deep Convolutional Networks for Large-Scale Image Recognition ». arXiv, 10 avril 2015. doi: 10.48550/arXiv.1409.1556.

[7]   A. Filonenko, L. Kurnianggoro, et K.-H. Jo, « Comparative study of modern convolutional neural networks for smoke detection on image data », in *2017 10th International Conference on Human System Interactions (HSI)*, juill. 2017, p. 64-68. doi: 10.1109/HSI.2017.8004998.

[8]   C. Nebauer, « Evaluation of convolutional neural networks for visual recognition », *IEEE Transactions on Neural Networks*, vol. 9, n° 4, p. 685-696, juill. 1998, doi: 10.1109/72.701181.

[9]   G. Litjens *et al.*, « A Survey on Deep Learning in Medical Image Analysis », *Medical Image Analysis*, vol. 42, p. 60-88, déc. 2017, doi: 10.1016/j.media.2017.07.005.

[10]   D. Dai, « An Introduction of CNN: Models and Training on Neural Network Models », in *2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR)*, nov. 2021, p. 135-138. doi: 10.1109/ICBAR55169.2021.00037.

[11]   R. Ramteke et M. Khachane, « Automatic Medical Image Classification and Abnormality Detection Using K- Nearest Neighbour », *International Journal of Advanced Computer Research*, vol. Volume-2 Number-4, p. 190-196, déc. 2012.

[12]   Khushboo Singh, Satya Verma, « Detecting Brain Mri Anomalies by Using Svm Classification », vol. 2.

[13]   S. Jayade, D. T. Ingole, et M. D. Ingole, « MRI Brain Tumor Classification using Hybrid Classifier », in *2019 International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET)*, déc. 2019, p. 201-205. doi: 10.1109/ICITAET47105.2019.9170147.

[14]   G. Garg et R. Garg, « Brain Tumor Detection and Classification based on Hybrid Ensemble Classifier ». arXiv, 1 janvier 2021. doi: 10.48550/arXiv.2101.00216.

[15]   R. Ramteke et M. Khachane, « Automatic Medical Image Classification and Abnormality Detection Using K- Nearest Neighbour », *International Journal of Advanced Computer Research*, vol. Volume-2 Number-4, p. 186 192, déc. 2012.

[16]   M. Tan et Q. V. Le, « EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks ». arXiv, 11 septembre 2020. doi: 10.48550/arXiv.1905.11946.

[17]   S. B. Patel, P. H. Patel, V. D. Jain, et J. P. Verma, « Improvised VGG16 CNN Architecture for Predicting Tuberculosis Using the Frontal Chest X-Ray Images », in *Smart Systems: Innovations in Computing*, Singapore, 2022, p. 69-80. doi: 10.1007/978-981-16-2877-1_7.

[18]   A. Krishnaswamy Rangarajan et R. Purushothaman, « Disease Classification in Eggplant Using Pre-trained VGG16 and MSVM », *Sci Rep*, vol. 10, n° 1, p. 2322, févr. 2020, doi: 10.1038/s41598-020-59108-x.

[19]   J. Su et H. Wang, « Fine-Tuning and Efficient VGG16 Transfer Learning Fault Diagnosis Method for Rolling Bearing », in *Proceedings of IncoME-VI and TEPEN 2021*, Cham, 2023, p. 453-461. doi: 10.1007/978-3-030-99075-6_37.

[20]   M.-S. Chen, C. Hwang, et T.-Y. Ho, « Terrain Image Classification with SVM », in *Advances in Swarm Intelligence*, Berlin, Heidelberg, 2013, p. 89-97. doi: 10.1007/978-3-642-38715-9_11.

[21]   X. Zhang, « Support Vector Machines », in *Encyclopedia of Machine Learning*, C. Sammut et G. I. Webb, Éd. Boston, MA: Springer US, 2010, p. 941-946. doi: 10.1007/978-0-387-30164-8_804.

[22]   Y. Lee, « Support Vector Machines for Classification: A Statistical Portrait », *Methods in molecular biology (Clifton, N.J.)*, vol. 620, p. 347-68, janv. 2010, doi: 10.1007/978-1-60761-580-4_11.

[23]   C. Sammut et G. I. Webb, Éd., « Logistic Regression », in Encyclopedia of Machine Learning, Boston, MA: Springer US, 2010, p. 631-631. doi: 10.1007/978-0-387-30164-8_493.

[24]   B. Srinivas, G. Sasibhushana Rao, « A Hybrid CNN-KNN Model for MRI brain Tumor Classification », vol. 8, doi: 10.35940/ijrte.B1051.078219.