Medical & Biological Engineering & Computing (2020) 58:1751–1765
https://doi.org/10.1007/s11517-020-02198-6

ORIGINAL ARTICLE

Machine learning–based identification of radiofrequency electromagnetic radiation (RF-EMR) effect on brain morphology: a preliminary study

Ritesh Maurya[1] · Neha Singh[2] · Tanu Jindal[2] · Vinay Kumar Pathak[3] · Malay Kishore Dutta[1]

Received: 15 May 2019 / Accepted: 22 May 2020 / Published online: 1 June 2020
© International Federation for Medical and Biological Engineering 2020

Abstract

The brain of a human and other organisms is affected by the electromagnetic field (EMF) radiations, emanating from the cell phones and mobile towers. Prolonged exposure to EMF radiations may cause neurological changes in the brain, which in turn may bring chemical as well as morphological changes in the brain. Conventionally, the identification of EMF radiation effect on the brain is performed using cellular-level analysis. In the present work, an automatic image processing–based approach is used where geometric features extracted from the segmented brain region has been analyzed for identifying the effect of EMF radiation on the morphology of a brain, using drosophila as a specimen. Genetic algorithm–based evolutionary feature selection algorithm has been used to select an optimal set of geometrical features, which, when fed to the machine learning classifiers, result in their optimal performance. The best classification accuracy has been obtained with the neural network with an optimally selected subset of geometrical features. A statistical test has also been performed to prove that the increase in the performance of classifier post-feature selection is statistically significant. This machine learning–based study indicates that there exists discrimination between the microscopic brain images of the EMF-exposed drosophila and non-exposed drosophila.

Keywords *Drosophila melanogaster* · Brain morphology · Radiofrequency electromagnetic radiation (RF-EMR) · Center-initialized *K*-means algorithm · Geometrical features · Genetic algorithm · Classifier

1 Introduction

Over the last couple of years, the increase in the number of telephone subscribers is many-fold. According to the reports of CNBC (Consumer News and Business Channel), the growth of mobile phone sales in percentage is increasing at a rapid pace. Digital 2019 report has mentioned that there are 5.11 billion unique mobile users in the world in January 2019, which is 100 million (2%) greater than the number of mobile users in January 2018. Thus, with the increase in the number of mobile phone users, it is highly probable that human and another organism will have more exposure to radiofrequency electromagnetic radiations (RF-EMR), emanating from mobile towers and mobile phones.

Several research studies have been conducted to analyze the effect of RF-EMR on the brain and other organs of humans as well as other organisms. It has been reported by some of the researchers that the exposure to these radiations has harmful effects on the brain. It has been found in the study that an hour of mobile phone use per day can significantly increase the risk of tumor [1] after 10 or more years [2, 3]. It may result in

✉ Malay Kishore Dutta
mkd@cas.res.in

Ritesh Maurya
ritesh@cas.res.in

Neha Singh
neha.singh.0574@gmail.com

Tanu Jindal
tjindal@amity.edu

Vinay Kumar Pathak
vinay@vpathak.in

[1] Centre for Advanced Studies, Dr. A.P.J. Abdul Kalam Technical University, New Campus, Lucknow 226031, India

[2] Amity Institute for Environmental Toxicology, Safety and Management, Amity University, Noida, India

[3] Dr. A.P.J. Abdul Kalam Technical University, Lucknow 226031, India

neurological, chemical, morphological, and electrical changes in the brain [4].

Mobile phones remain in close proximity with the brain during talk mode [5] and radiofrequency electromagnetic fields can penetrate 4–6 cm into the human brain [6]. Various investigations [7, 8] have shown effects of RF-EMR on the development and the morphology of different organs in rats. It was studied and observed the adverse effects of 900 MHz RF-EMR [9] on the development of hippocampus, morphology, and learned response in Wistar rats [10, 11]. Various studies conducted between 2007 and 2019 have reported that RF-EMR affected the activities [12, 13]/morphology [14–16]/chemistry [17, 18] of the hippocampus in animals.

The changes in the morphology of a brain occur due to cell death or DNA damage at the cellular level [19]. Apoptosis or programmed cell death is an important biological event in tumor promotion, which may be enhanced due to DNA fragmentation [20]. These authors have also reported an increased level of DNA strand break (single and double) due to microwave exposure in rat brain at various frequencies (2.45 and 50 GHz) [21, 22]. The experiment, which is used to conduct, for the identification of EMF radiation on the brain at the cellular level, is expensive and a time consuming one. The experiment requires the preparation of cell suspension for further analysis and the preparation of cell suspension usually takes 1 to 2 days, when the proper procedure of protocols is followed. A lot of sample cells are also lost during the treatment of the cells which varies the experimental results significantly. Another issue with the handling of drosophila brain is that being a very tiny organ, it is very difficult to handle it and one has to be very careful while working with it. Even a difference in microliters, of cell suspension treatment chemical, can affect the experimental results significantly.

In the proposed work, machine learning and computer vision methods have been used to analyze the changes in the morphology of the brain due to the effect of RF-EMR. Drosophila melanogaster has been used as a specimen for the experimentation purpose since drosophila is known to be a well suited and highly tractable genetic model organism for an understanding of the biological, physiological, and neurological mechanism in the body of a human [23].

The changes in the morphology of a drosophila brain when exposed to the EMF radiations can be correlated with the changes that took place in the brain at the cellular level [24]. Various geometry-based features have been extracted (from the segmented microscopic brain image) and analyzed, to detect the morphological changes that took place in the drosophila brain due to its exposure to EMF radiations. The machine learning classifier being learned on those features might be capable of correlating these morphological changes in the drosophila brain with their geometrical features. The contributions of the present work can be summarized as follows:

i. The main contribution of this work is in the development of a machine learning and computer vision–based automated system that can be used to predict effects of EMF radiations emanated from mobile stations in the brain by using *Drosophila melanogaster* as a specimen.
ii. Another contribution of this paper lies with the segmentation of brain region from the input microscopic image (using center-initialized *K*-means algorithm) with a Dice similarity coefficient of 0.9611.
iii. Certain geometric features have been determined and selected from a set of extracted features when they are used to train any machine learning classifier; machine learning classifier establishes discrimination between the drosophila brain images in EMF-exposed and sham-exposed class.
iv. The fitness function of the genetic algorithm has been designed in such that it provides a tradeoff between the accuracy of a classifier and the size of the selected subset of features.

The rest of the paper is organized in the following manner: Section 2 of the paper explains the materials used in the present work. Section 3 discusses the methods used in the present work. Section 4 reports and discusses the experimental results. Section 5 made a conclusive remark about the finding of the present work.

2 Materials

2.1 Sample preparation

Wild type red-eyed Oregon-R strain of drosophila flies was obtained from Drosophila Stock Centre, Mysore University, India, and was cultured and maintained in the insect rearing lab Amity University, Noida. The flies were kept at room temperature environment at 25 ± 1 °C and relative humidity of 70% with 12-h periods of light and dark periods on standard Drosophila media and were kept to obtain generations so as to have a pure strain of the organism. For maintaining Drosophila cultures, most stocks were successfully cultured by periodic mass transfer of adults to fresh food.

RF-EMR exposure was performed by using a horn antenna at 2400 MHz of installed frequency. The RF-EMR source produced 1800 to 2700 MHz frequency with an antenna gain of 10 db (relative). The computed power density of the RF-EMR exposure to carry out this experiment was 29 mW/m^2. The flow diagram for the exposure chamber has been shown in Fig. 1.

The exposure chamber during the experiments was maintained at a temperature of 25 °C ± 0.1 °C to avoid thermal effects. The chamber is also lined by anechoic material to prevent any reflected radiation waves which can interfere with

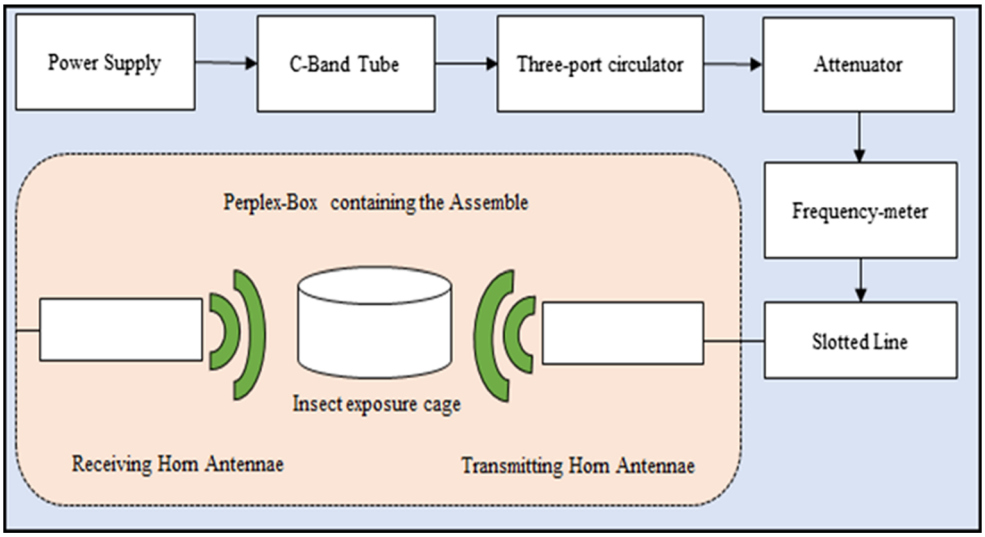


Fig. 1 Flow diagram for exposure chamber

the experiment. For the RF-EMR exposure around 100, 5-day-old adult flies were taken in an exposure tube made up of acrylic material having dimensions of 8 × 3 × 3 cm. A wet filter paper is kept at the bottom to maintain the humidity inside the tube. The experiments are carried out in two sets of sham (without) exposed and exposed groups. The sham-exposed group flies were kept in the same conditions of exposure (in the same chamber, temperature, and humidity) but without exposure and exposed group flies were exposed to RF-EMR. Drosophila in EMF-exposed group was exposed to RF-EMR for 12 h per day continuously for 5 days at same time of the day.

2.2 Image acquisition

After exposure, the flies were sacrificed and their brains were dissected under stereo-microscope. The whole dissection is performed in phosphate-buffered saline (PBS). The images of the brain were taken in all the four directions by using Magcam DC 5 (5.1MP) fitted along with the microscope. The resolution of a microscopic image of the brain of drosophila is 1944 × 2592 pixels. Fig. 2 shows an image acquisition setup used to capture the image of the drosophila brain.

The dataset used for the present study consists of microscopic images of drosophila brain in two sets, i.e., RF-EMR-exposed and sham-exposed. The total number of microscopic images used for the present study in both groups, i.e., RF-EMR-exposed and sham-exposed (control) has been described in Table 1.

2.3 Analysis of the acquired image samples

The effect of RF-EMR on drosophila brain has been performed by analyzing the color microscopic image of drosophila brain. Fig. 3 highlights the reflection and noise that occurred during the acquisition of microscopic images of the brain and also displays random samples of microscopic images for each class, i.e., RF-EMR exposed and sham-exposed.

No visual difference can be observed between the brain images in two groups. Hence, no morphological discrimination can be established just by the visual analysis of the brain images. Therefore, to analyze the morphological differences between the brain images of both groups, the brain region has to be segmented from the acquired microscopic images and the detailed analysis of morphological features extracted from



Fig. 2 Image acquisition setup

Table 1 Description of the dataset

Class/group	Total
Sham exposed (control)	79
Exposed (treated)	79
Total	158

the microscopic brain images present in both groups has been made.

3 Methods

Microscopic brain images of drosophila in EMF-exposed and sham-exposed classes were acquired. The captured images are used as an input to the methodology proposed in the present work. The proposed methodology for the present work can be divided into the following steps: first step consists of the acquisition of microscopic images of drosophila brain. The second step deals with the pre-processing and segmentation of brain region from the pre-processed image. In the third step, geometrical features have been extracted from the segmented brain region. In the fourth step, the discriminatory subset of features is selected by using the genetic algorithm–based feature selection algorithm. In the fifth and the last step, the selected features are used to train and test the machine learning models, for identification of morphological difference, if present, in the brain images of both the groups. The methodology proposed for the present work has been shown in Fig. 4.

3.1 Brain region segmentation

There appear reflection and other artifacts in the acquired microscopic images, which makes the segmentation process difficult. These artifacts and reflections are basically the noise which disrupts the quality of acquired brain image. Therefore, it is necessary to pre-process the original microscopic image in an appropriate color model prior to segmentation. An input image has been analyzed in different color models such as RGB, HSV, and YCbCr [25, 26]. Among these color models, the brain region is distinctly visible in the luminance (Y) component of the YCbCr model; therefore, pre-processing has been performed on the luminance component of the YCbCr model. The pre-processed Y image is used for segmentation of brain region using the proposed center-initialized *K*-means algorithm–based segmentation procedure.

3.1.1 Pre-processing

Since the original image captured with the stereo-microscope has been corrupted with the noise, therefore the Y-component image of YCbCr model is filtered with the median filter. The contrast of median filtered image is adjusted by the intensity adjustment technique [25] to highlight the region of interest.

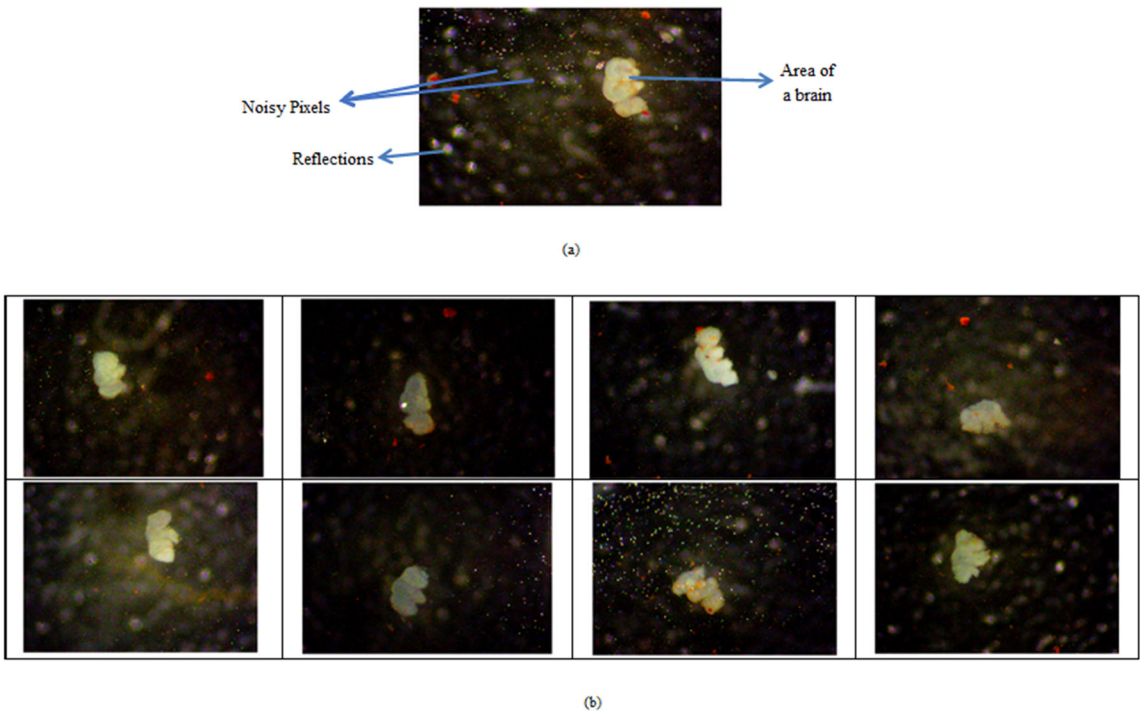


Fig. 3 Microscopic brain image. (a) Description of noise and reflection present in the sample. (b) RF-EMR exposed group brain images (row 1) and sham-exposed brain images (row 2)

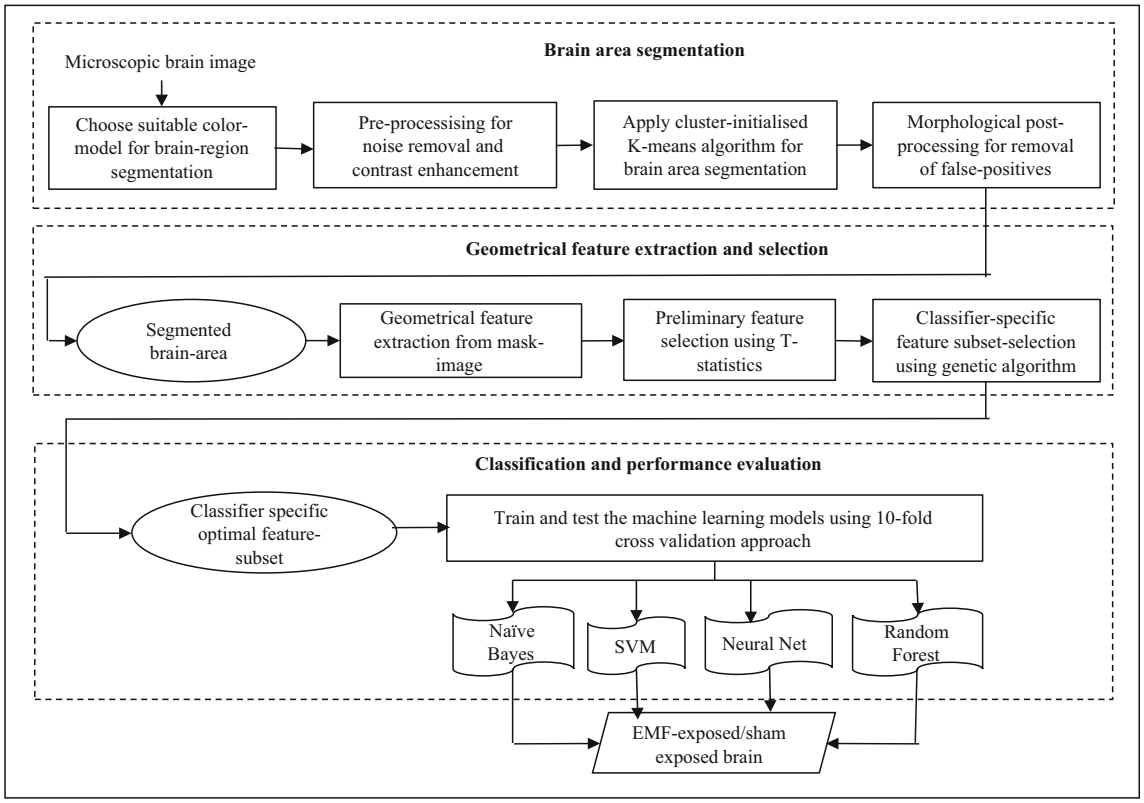


Fig. 4 The proposed methodology

Fig. 5 shows the original image, Y-component filtered image, and the contrast-stretched filtered image.

3.1.2 Segmentation procedure

The intensity of pixel present in the region of the brain (as shown in Fig. 5(c)) is almost similar. Therefore, the region of the brain can be segmented from the original image by creating a cluster of pixels present in the brain region. But for the pixels present in the brain region to form a single cluster, the appropriate value of initial cluster centers, as well as the number of clusters, should be chosen. Fig. 6 describes the detailed procedure involved segmentation process of the brain.

The choice of an appropriate value of initial cluster centers is equally important as with a large number of cluster centers (K) results in over-segmentation, the opposite is also true for the smaller number of cluster centers. For the present problem of brain region segmentation, the $K > 3$ over segments whereas $K < 3$ under segments the input image. Therefore, the value of $K = 3$ has been chosen for the segmentation of the brain correctly.

The initialization of cluster centers in the K-means algorithm is also very important [27, 28] as random initialization of cluster center results in wrong segmentation of the region of interest. In the present work, three cluster centers (c_1, c_2, and c_3) are chosen in such a way that it divides the overall gray level range into four parts d_1, d_2, d_3, and d_4 of equal size. Initialization of cluster centers using the proposed method helps in faster convergence of the proposed segmentation algorithm as brain region pixels form a cluster around the cluster center c_3 as the algorithm proceeds.

The input to the proposed segmentation algorithm is the pre-processed Y image. The procedure for the proposed segmentation algorithm has been described as follows:



Fig. 5 Pre-processing steps. (a) Original image. (b) Median filtered luminance (Y) image. (c) Contrast-enhanced image

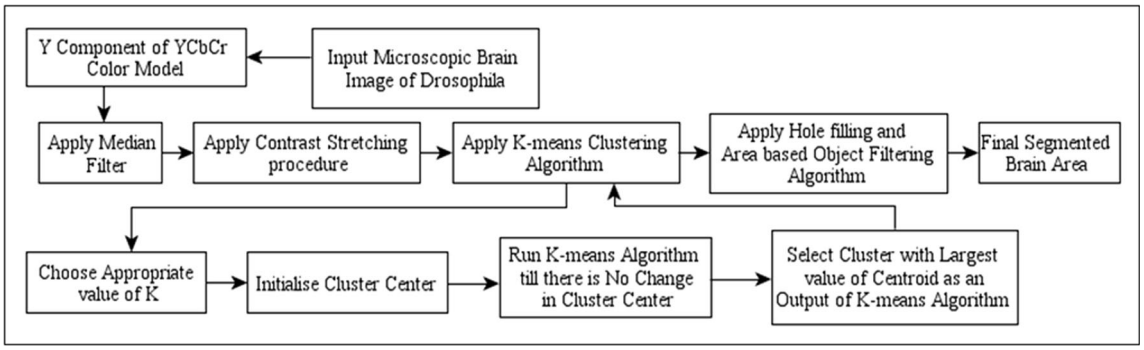


Fig. 6 Detailed description of the proposed segmentation algorithm

Proposed segmentation algorithm:

i. Initialize the number of cluster centers (K) and initialize their values.
ii. Calculate the Euclidean distance between the cluster centers and all the other pixels in an image.

$$d_{x,y} = f(x,y) - c_k \tag{1}$$

where $d_{(x,y)}$ is the distance of pixel at coordinates x and y from the kth cluster center c_k

i. Assign pixel to nearest cluster center based its Euclidean distance d from cluster center c_k.
ii. After all, pixels are assigned to their nearest center, recalculate the cluster center using the relation mentioned below:

$$c_k = \left[\frac{\sum_{i=1}^{n} f_i(x,y)}{n}\right] \text{ where } (x_k, y_k) \in c_k \tag{2}$$

where n is the total number of pixels in cluster centered at c_k. Repeat steps 2, 3, and 4, till there is no change in the value of the cluster center.

iii. Morphological post-process the segment formed for the region of the brain.

The output of the proposed algorithm is the segmented brain region. The intermediate (step-wise) outputs of the proposed segmentation algorithm have been shown in Fig. 7.

After segmentation of the brain region from the colored microscopic image, the geometrical features are extracted from the segmented binary mask image of the brain to analyze its morphological features. The feature extraction process used in the present work has been described in the next section.

3.2 Feature extraction

To analyze the effect of RF-EMR on drosophila brain, a certain set of geometrical features [26] has been extracted from the segmented binary mask of the brain region, represented as I_B. The reason for choosing these 11 geometrical features have been explained as follows:

i. Area: The area has been chosen for the analysis of the morphology of drosophila brain since the area of an object is indicative of its relative size. Thus, any kind of morphological change in the brain is likely to impact its area.

$$\text{Area} = \sum_{n=1}^{n=j} \sum_{m=1}^{m=j} \left(I_{B_{(m,n)}}\right) \tag{3}$$

where $I_{B_{(m,n)}}$ is the pixel of an image I_B at a location (m, n).

ii. Major axis length: The major axis length has been chosen for the analysis of the morphology of drosophila brain in the present work, since it is the length of the longest line connecting the two points on the perimeter of a brain. Thus, if the morphology of a brain undergoes change, then it might also bring changes in the major axis length of a brain. Major axis length can be computed by the equation given below:

$$\text{MajorAxisLength} = \text{length}(\text{MajorAxis}(I_B)) \tag{4}$$

iii. Minor axis length: The minor axis length has been chosen for the analysis of the morphology of drosophila brain since it is the maximum distance between two border points located on the perimeter of a located in a binary image I_B, in the direction, perpendicular to the major axis length.

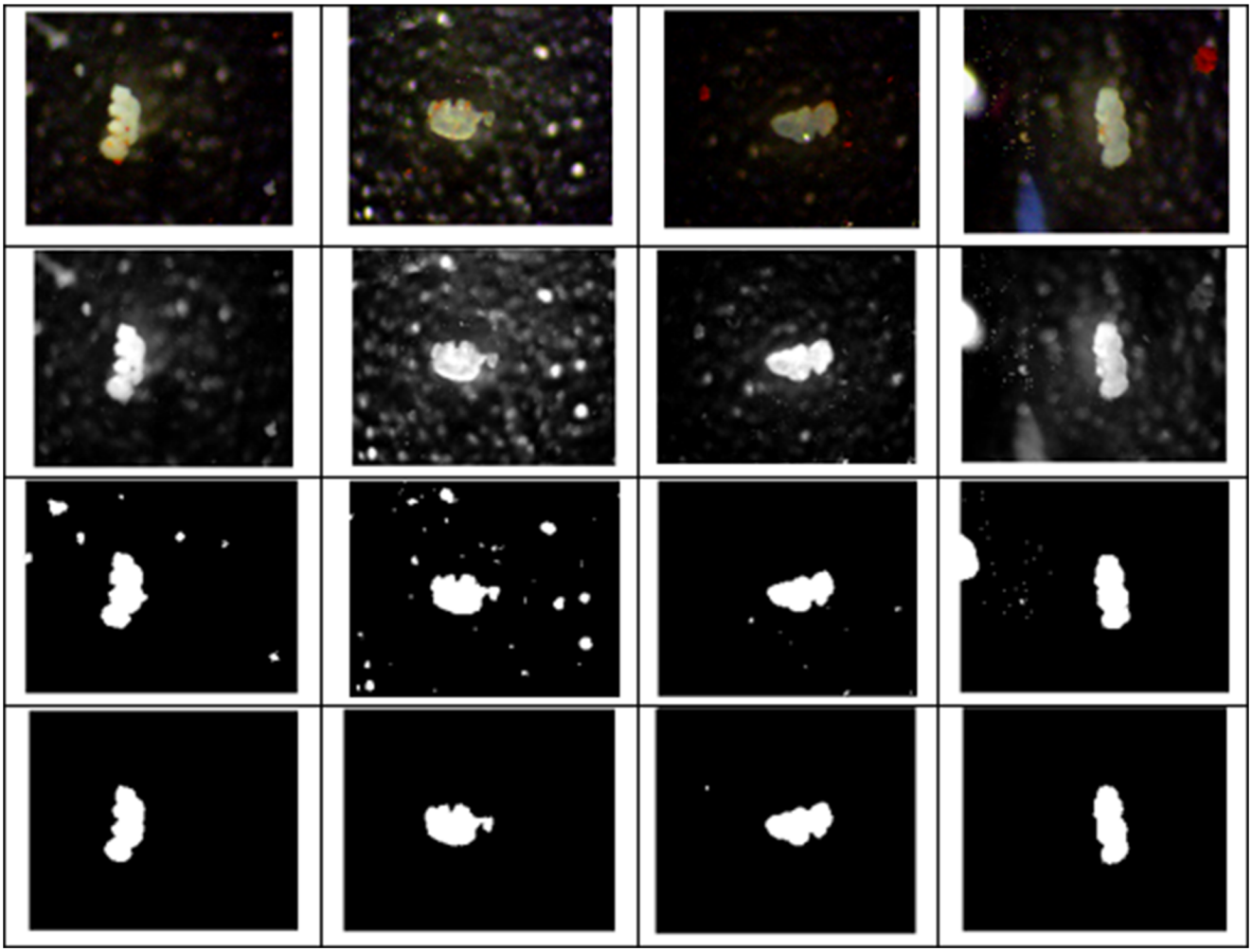


Fig. 7 Intermediate to final outputs in segmentation process. (a) The original microscopic image. (b) Pre-processed luminance (Y) image. (c) Cluster formed by the brain region pixels. (d) Binary mask of the brain region

$$\text{MinorAxisLength} = \text{length}(\text{MinorAxis}(I_{\text{B}})) \tag{5}$$

iv. Eccentricity: The eccentricity has been chosen for the analysis of the morphology of the drosophila brain since it is the measure of the deviation of elliptical path. Thus, if the morphology of a brain change, then the deviation in the elliptical curve of a brain boundary is expected to change.

It is the ratio of the distance between the foci of the object and its major axis length. The mathematical formula to calculate eccentricity is:

$$\text{Eccentricity} = (\text{Distance between foci}/\text{major axis length}) \tag{6}$$

v. Convex area: Total number of pixels in the convex image of an object is known as convex area. Thus, if any morphological changes occur in the brain, then there is a possibility that the convex area of a brain might have changed.

vi. Compactness: It is a quantitative measure to represent the degree to which an object is compact. It is calculated using the equation given below:

$$\text{Compactness} = \frac{P^2}{4\pi A} \tag{7}$$

where P is the perimeter of a brain and A is the area of a brain region.

vii. Euler number: It describes the relationship between the number of contiguous parts and number of holes

in an image. Any morphological changes that may occur in the brain are likely to impact the Euler number of the brain.

If N_c is the total number of contiguous regions in an image and N_h is the total number of holes in an object, then Euler number can be calculated as follows:

$$\text{Euler number} = N_c - N_h \tag{8}$$

viii. Equiv diameter: It represents the diameter of a circle whose area is equal to the area of an object. Any morphology deviation of a brain due to EMF exposure is likely to affect the equiv diameter of a brain. It can be calculated as follows:

$$\text{EquivDiameter} = \sqrt{4*\frac{\text{Area}}{\pi}} \tag{9}$$

ix. Solidity: Solidity is the measure of convexity or concavity of an object. For current problem, it is the ratio of the area of a brain region and the area of the convex hull, which is the smallest convex polygon containing all the pixels in the brain region and is likely to have discriminatory behavior under the effects of EMF exposure.

$$\text{Solidity} = \frac{\text{Area}_0}{\text{Area}_c} \tag{10}$$

where Area_o is the area of a segmented image and Area_c is the area of a convex hull surrounding the segmented brain image in I_B.

x. Extent: Extent defines the ratio of the total number of non-zero pixels in an object and the total number of pixels in the corresponding bounding box. Thus, if any morphological changes occur in the brain, then the extent feature is likely to undergo changes.

$$\text{Extent} = \frac{N_o}{N_b} \tag{11}$$

where N_o total number of pixels in an object and N_b is the total number of pixels in the corresponding bounding box.

xi. Perimeter: Perimeter has been chosen for analysis, since it is a distance around the boundary of a segmented brain area in I_B. Morphological changes that may occur in the brain are likely to affect the perimeter of the brain.

A visual analysis of the features extracted from RF-EMR and the sham-exposed group has also been shown in the box plot shown in Fig. 8.

All the features described above have some relevance to the geometry of the brain but not all the features have the equal discriminative ability. Therefore, to select the highly discriminatory features among all the extracted features, genetic algorithm–based feature selection has been used and described in the next section.

3.3 Discriminatory feature selection

To select a set of features which establishes the morphological discrimination between the brain images of the two groups, proper feature selection procedure needs to be adopted. Basically, feature selection methods fall into two categories: filter method and wrapper method where filter methods are independent of the machine learning models to be learned and are computationally cheaper, whereas the wrapper methods select set of features after evaluating the performance of a classifier on those features [29]. The advantage of both the methods has been utilized to select the classifier specific optimal subset of features, in the present work.

At first, the selection of features is performed on the basis of *T*-statistics scores of the extracted features; it removes the indistinct features in the initial phase of the feature selection process. *T*-statistics is a supervised feature selection approach [30]. It assesses whether the mean for a feature in two classes (exposed/sham-exposed) is statistically different from each other. Each sample of class 1 is labelled as 0 and for class 2, as 1. Then, for each feature f_i, the mean μ_i^0 (resp. μ_i^1) and standard deviation δ_i^0 (resp. δ_i^1) are calculated using only the samples labelled 0 (resp. 1). Then, a score $T(f_i)$ can be calculated using Eq. (1) given below:

$$T(f_i) = \frac{\mu_i^0 - \mu_i^1}{\sqrt{\frac{\left(\delta_j^0\right)^2}{n_0} + \frac{\left(\delta_j^1\right)^2}{n_1}}} \tag{12}$$

where n_0 and n_1 represents the number of samples labelled as 0 (resp. 1). The higher the value of *T*-score for a particular geometric feature, the higher its discriminatory behavior is.

The scores for all the features are represented in a stem graph as shown in Fig. 9. It can be observed from Fig. 9 that the features numbered 2 and 4 are the most discriminatory, whereas the features numbered 5 and 10 are the least

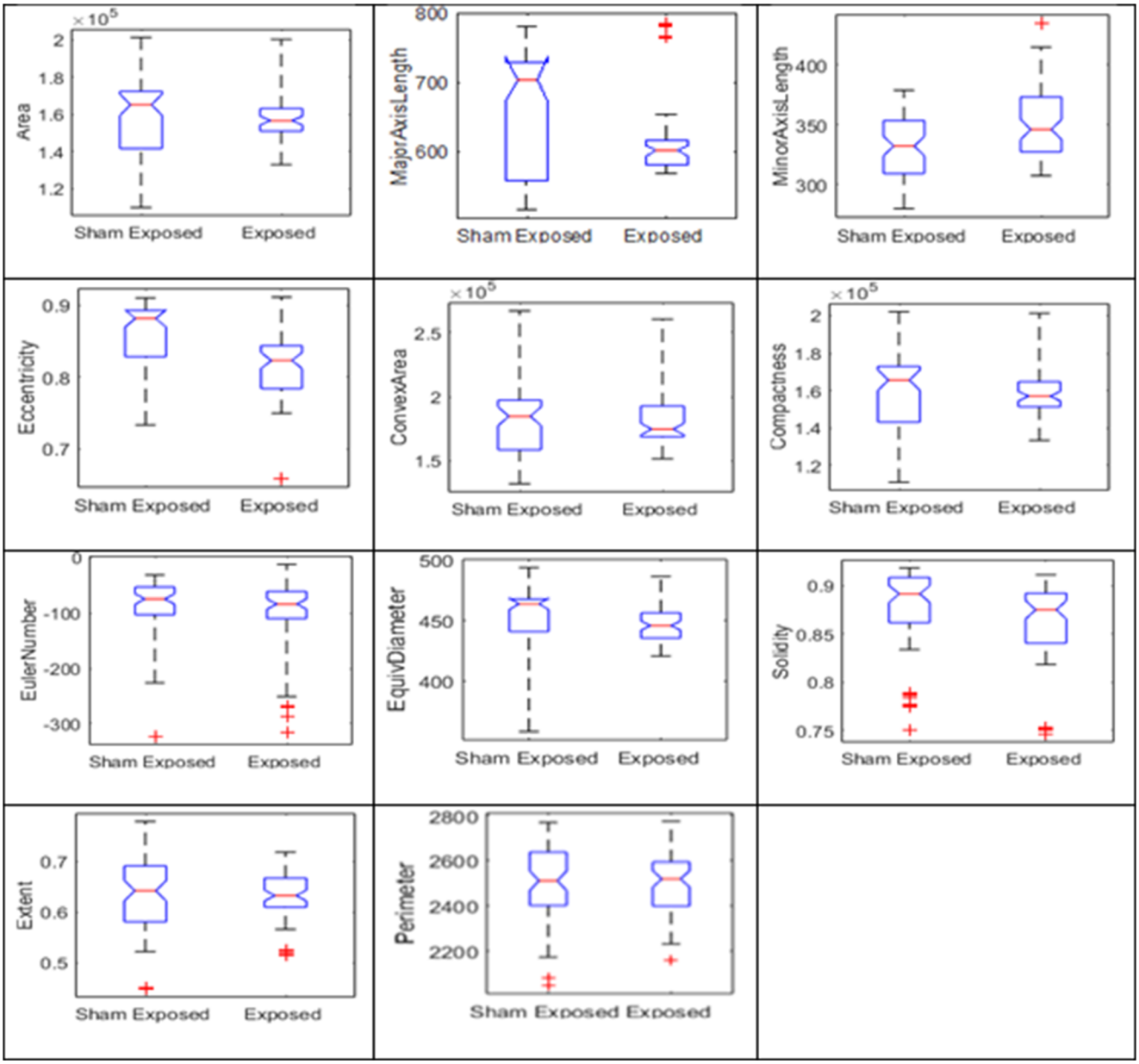


Fig. 8 Box plot comparison of features extracted from the brain images in RF-EMR exposed and sham-exposed class

discriminatory. With a threshold value of 0.5285, a total of nine (1, 2, 3, 4, 6, 7, 8, 9, 11) features have been selected in preliminary feature selection step, where the features numbered as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11 correspond to the features area, major axis length, minor axis length, eccentricity, convex area, compactness, Euler number, equiv

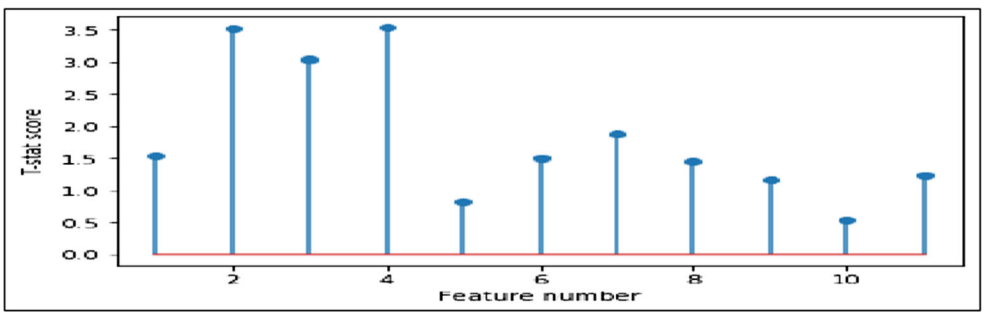


Fig. 9 *T*-stat score of the extracted features

Table 2 Hyperparameters of genetic algorithm

Parameter/method	Value/type
Population size	8–12
Number of generations	10
Crossover probability	0.7
Mutation probability	0.06
Parent selection procedure	Roulette wheel
Crossover	Two-point
Mutation	Flip-bit
Number of offspring	8
Elitism	2

diameter, solidity, extent, and perimeter of the segmented brain region respectively.

Evolutionary methods have been used for feature selection and received much success in the recent past [31]. Therefore, after removal of non-discriminatory features in the initial phase using the *T*-statistics feature selection method, the rest of the remaining features are fed into the genetic algorithm, to select optimal subset of features among 2^9 possible combinations of subsets. The optimal subset of features has been found for the Naïve Bayes, support vector machine (SVM), artificial neural network (ANN), and random forest classifiers. The fitness function of genetic algorithm for selection of any *i*th chromosome is given by Eq. 2.

$$\text{fit}_i = w_1 * \text{accuracy}(z_i) + w_2 * \frac{1}{r_f} \tag{13}$$

where w_1 and w_2 are the weights used in fitness function which lies between 0 and 1 and $w_2 = w_1 - 1$ and r_f is the ratio of the size of selected feature subset (no. of 1s in chromosome) and the total number of features, i.e., eleven; fit_i is a fitness function of *i*th selected subset of features represented by z_i for a given classifier.

The first component of the genetic algorithm fitness function represents weighted accuracy (10-fold cross-validated) of a classifier on *i*th selected feature subset (chromosome) represented by z_i and the second component represents the weighted ratio of length of selected feature subset (total no. of 1s in chromosome) and the size of features left, after application of *T*-statistics filter based selection method. The weight values w_1 and w_2 provide a tradeoff between the accuracy of classifier on selected features and the size of selected subset of feature. Value of $w_1 = 0.8$ and $w_2 = 0.2$ has been used for the present work as these values of weights provide an optimal tradeoff between accuracy and the size of selected set of features.

Steps performed by genetic algorithm for feature selection are described as follows:

Step 1. In the first step, each feature subset (chromosome) is encoded by the binary string of 0s and 1s, where each bit represents presence (if, 1) or absence (if, 0) of individual feature in the chromosome.

Table 3 List of features selected by the genetic algorithm

S.No.	Feature	Classifiers			
		Naïve Bayes	SVM	ANN	Random forest
1	Area	✗	✗	✓	✗
2	Major axis length	✓	✓	✓	✓
3	Minor axis length	✗	✗	✗	✗
4	Eccentricity	✗	✗	✓	✗
5	Convex area	----------Removed using *T*-statistics method----------			
6	Compactness	✓	✗	✓	✗
7	Euler number	✗	✓	✗	✗
8	Equiv diameter	✓	✗	✗	✗
9	Solidity	✓	✗	✓	✓
10	Extent	----------Removed using *T*-statistics method----------			
11	Perimeter	✓	✓	✓	✗
Total number of selected features		5	3	6	2

Table 4 Hyperparameters of machine learning classifiers

Machine learning classifier	Hyperparameters
Naïve Bayes	Gaussian distribution
Support vector machine	Gaussian radial basis function kernel, scaling factor sigma = 1.8
Artificial neural network	No. of hidden layers = 5, training function = Levenberg-Marquardt, Epochs = 30
Random forest	Number of trees = 5

Table 5 Performance metrics of classifiers on selected features

	Classifiers			
	Naïve Bayes	Support vector machine	Artificial neural network	Random forest
Sensitivity	91.9%	88.57%	98.6%	98.55%
Specificity	81.70%	86.48%	97.26%	94.67%
F1 score	0.86	0.87	0.97	0.96
Accuracy	86.11%	87.50%	97.91%	96.52%
Youden's index	0.74	0.76	0.96	0.93

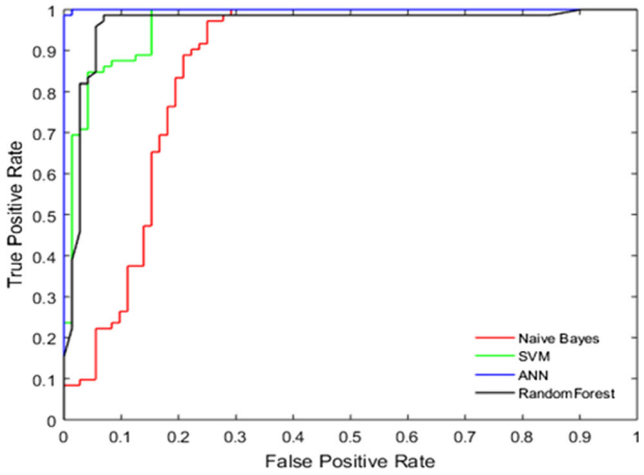


Fig. 10 ROC curve for classifier

Step 2. Start the algorithm with a randomly generated population of chromosomes. The size of the population varies from 8 to 12.

Step 3. Fitness of each individual chromosome in population is calculated by the fitness function given in Eq. 2.

Step 4. Select parents using the roulette wheel selection method (also named proportional selection) from the population in the current generation. In roulette wheel selection, probability of selection is proportional to the fitness value of chromosomes.

Step 5. Use two-point crossover and mutation with the given probability of crossover and mutation in order to generate offspring in the next generation.

Step 6. Stop if the fitness function of chromosomes does not change for successive five generations. Select the fittest chromosome from the population as a solution to a current problem; otherwise, go to step 3.

The output of the algorithm is the feature selected by the genetic algorithm. The parameters and methods used for feature selection using a genetic algorithm are shown in Table 2.

The list of features selected by the genetic algorithm has been shown in Table 3. The "✓" symbol represents the feature has been selected and the "✗" symbol represents the feature has been removed from the list.

The features selected by the proposed feature selection method were given as input to the machine learning classifiers to test whether any discrimination is present between the features of both the classifiers. The details of machine learning classifiers are given in the next section.

3.4 Machine learning for classification

Machine learning model plays an important role in establishing a discrimination between RF-EMR-exposed and sham-exposed groups on the basis of classification of segmented brain images in the test set into their respective classes. Four different types of classifiers such as Naïve Bayes, support vector machine (SVM), artificial neural network (ANN), and random forest have been used in this work for classification purpose. Discriminatory features selected using a genetic algorithm–based feature selection algorithm have been fed into the machine learning classifiers to train and test the performance of features on the selected features. The hyperparameter settings for four classifiers used in the present work has been shown in Table 4.

4 Experimental results and discussion

To test the discrimination between the drosophila brain images in both the classes, the performance of the classifiers on the features selected by the genetic algorithm has been analyzed. A comparative analysis has also been performed on the performance of classifiers, trained with and without selected features. The performance of the classifiers has been validated using 10-fold cross-validation. All the algorithms presented in this work have been implemented on CPU@2.0GHz, 4 GB RAM, 64-bit operating system. The performance of classifiers has been

Table 6 Comparison chart for classifier performance before and after feature selection

Classifier	Average % accuracy		Total no. of features		An absolute reduction in no. of features	%Increase in accuracy
	Before feature selection	After feature selection	Before feature selection	After feature selection		
Naïve Bayes	81.19%	86.11%	11	5 [2,6,8,9,11]	6	6.059
SVM	85.72%	87.50%	11	3 [2, 7, 11]	8	2.076
Artificial neural network	97.13%	97.91%	11	6 [1,2,4,6,9-,11]	5	0.80
Random forest	95.64%	96.52%	11	2 [2,9]	9	0.92

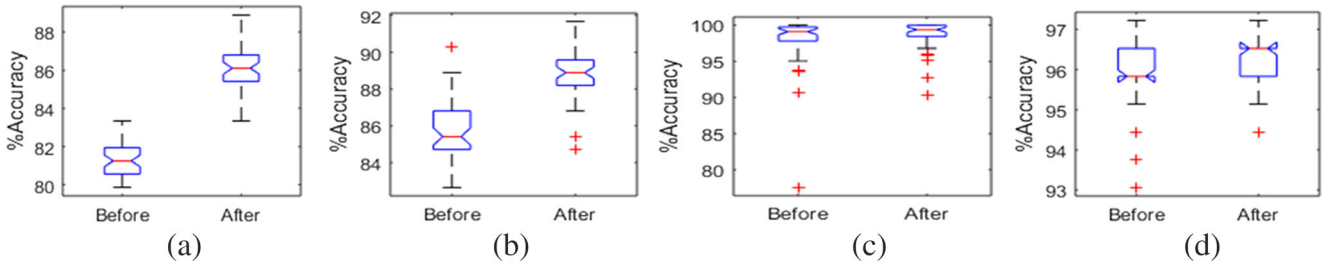


Fig. 11 Box plot representation for pre- and post-feature selection performance of the classifiers for 50 iterations. (a) For Naïve Bayes. (b) For SVM. (c) For ANN. (d) for random forest

evaluated on the parameters such as sensitivity, specificity, F1 score (harmonic mean of precision and recall), accuracy, and Youden's index (ROC curve which is farthest from the line of equality). The performance of the proposed segmentation algorithm has been measured with the Dice similarity coefficient which was evaluated to 0.9611 for the present segmentation problem. The average performance metrics of the classifiers (for10-fold) when trained with selected features have been shown in Table 5.

Fig. 10 shows the mean receiver operating characteristic (ROC) curve plot (for 10-fold) for all the four classifiers when trained with the selected features. Fig. 10 shows that the best performing classifier is neural network with an AUC value of 0.98, whereas Naïve Bayes performance was found to be worst with the AUC value of 0.89.

Table 6 presents a comparison chart showing the reduction in the number of features before and after feature selection as well as the corresponding increase in the average accuracy of a machine learning classifier for 10 folds.

Box plots representing the comparison between the performance of classifiers (for 50 iterations of 10-fold cross-validation) when trained with selected and all the eleven features have been shown in Fig. 11. Box plots shown in Fig. 11 show mean, standard deviation, and minimum and maximum accuracy for a classifier trained with or without feature selection.

Table 7 CPU time required in image segmentation and feature extraction

Sample images	Segmentation of brain area (in seconds)	Geometrical feature extraction (in seconds)
Image 1	3.95	.2255
Image 2	4.55	.2090
Image 3	7.76	.1955
Image 4	4.68	.1988
Image 5	3.77	.2029
Image 6	5.56	.1856
Image 7	3.84	.1888
Image 8	3.43	.2091
Image 9	4.45	.1987
Image 10	3.56	.1873

The CPU time required for brain region segmentation from input microscopic images and also the time required for geometrical feature extraction for some randomly selected samples have been shown in Table 7.

The total average time to segment brain area from all the microscopic images in the dataset was 4.21 s, whereas it took an average time of .1938 s for extraction of geometrical features from the segmented brain images.

Table 8 represents the reduction in the training time of a classifier post-feature selection.

It can be observed from Table 8 that classifier accuracy gets increased post-feature selection. But it is not reliable to report that the improvement in the classifier performance post-feature selection only on the basis of average accuracy of a classifier since means are susceptible to outliers [32]. Thus, it might not result in the identification of discriminatory features which are actually distinct.

To prove the increase in classifier accuracy post-feature selection of features is statistically significant, paired T test has been performed. Statistical significance test proves that the increase in classifier accuracy is not just by chance but with a confidence level of 0.95. Statistical T test has been used to check the hypothesis that the increase in an average accuracy of a classifier post-feature selection is statistically significant. The closer the T-value is to 0, the similar the two groups are. The p value denotes the probability of the increase in the accuracy of a classifier post-feature selection. If the p value is equal to 0.05, then there is 5% probability that increase in the accuracy of classifier is by chance. The test results after performing a paired T test on the accuracy of classifiers pre- and post-selection of features have been shown in Table 9.

Table 8 Classifier training time required before and after feature selection

Classifier	Training time	
	Before feature selection	After feature selection
Naïve Bayes	5.2351	2.1130
SVM	4.2197	1.1540
ANN	5.3432	2.2289
Random forest	6.0599	2.3200

Table 9 Results of the paired *T* test

Classifier	h	*T*-value	*p* value	Confidence interval
Naïve Bayes	1	23.5391	2.3053e−28	[− 5.5475…..…..…..− 4.6748]
SVM	1	− 9.4939	1.0968e−12	[− 3.8538…..…….…..− 2.5073]
ANN	0	− 2.5850	0.2350	[− 0.5144…..…….…..0.3405]
Random forest	1	− 2.2604	0.0283	[− 0.6297…..…..…..…..− 0.0370]

In Table 1, $h = 1$ represents the rejection of the null hypothesis and when $h = 0$ represents the acceptance of null hypothesis. Null hypothesis states that there is no significant increase in classifier accuracy post-selection of features, and is accepted in case of ANN and it has been rejected in case of other classifiers used in the present work for classification.

The following observations can be made by correlating the performance of classifiers discussed in this section and the features selected by the proposed feature selection algorithm as shown in Table 3.

i. The results presented in this section indicates that the features such as major axis length, perimeter, and solidity are the most discriminatory as these features have been selected by most of the classifiers and thereby also improved the performance of classifiers when tested.
ii. Considering the stability in the performance of the classifier and the corresponding increase in the accuracy of the classifier, the random forest classifier identifies the major axis length and solidity as the most discriminatory geometrical features among other nine.
iii. Minor axis length, convex area, and extent can be considered redundant features since the removal of these features does not affect the performance of the classifiers.
iv. It has been observed that no clear distinction can be made between the test set images on the basis of their geometrical areas as the area has only been chosen as a discriminatory feature by only one classifier, i.e., ANN. Thus, it does not qualify for discriminating the test images based on the area chosen as a geometrical feature.
v. The training time of the classifier post-feature selection has been reduced significantly because of the significant reduction in feature space.
vi. All the observations presented in this work are based on the microscopic images of the brain in RF-EMR and sham-exposed class and they do not claim any kind of biological or physiological changes that took place in the brain due to RF-EMR exposure.

5 Conclusion

The current work presented a novel pilot study to analyze the effect of RF-EMR on the morphology of drosophila brain as well as provide a computer vision–based methodology using machine learning. Region of the brain is automatically segmented using center-initialized *K*-means algorithm with a Dice similarity coefficient of 0.9611. A total of 11 geometrical features have been extracted from the drosophila brain images in RF-EMR-exposed and sham-exposed classes and their discriminative capabilities for the present classification problem have been tested. Classifier specific optimal subset of features has been determined using genetic algorithm–based feature selection method. Artificial Neural Networks perform with an average classification accuracy of 97.91%, which is the highest among other classifiers used. The discrimination among the features has been established with their relationship to the performance of classifiers on the features selected by genetic algorithm. The study establishes discriminatory features of the microscopic brain image of the specimen using machine learning methods. Future work in this may involve the use of other specimens to be exposed and studied for similar effects of RF-EMR.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical clearance In our knowledge, as per the Indian laws, working with *Drosophila melanogaster* does not require ethical clearance.

References

1. Ouadah NS, Lecomte A, Robidel F, Olsson A, Deltour I, Schüz J, Blazy K, Villégier AS (2018) Possible effects of radiofrequency electromagnetic fields on in vivo C6 brain tumours in Wistar rats. J Neuro-Oncol 140(3):539–546
2. Morgan LL, Miller AB, Sasco A, Davis DL (2015) Mobile phone radiation causes brain tumors and should be classified as a probable human carcinogen (2A) (review). Int J Oncol 46(5):1865–1871
3. Khurana VG, Teo C, Kundi M, Hardell L, Carlberg M (2009) Cell phones and brain tumors: a review including the long-term epidemiologic data. Surg Neurol 72(3):205–214
4. Lai H (2012) Neurological effects of non-ionizing electromagnetic fields, BioInitiative Working Group
5. Mausset AL, de Seze R, Montpeyroux F, Privat A (2001) Effects of radiofrequency EMF exposure on the GABAergic system in the rat cerebellum: clues fromsemi-quantitative immunohistochemistry. Brain Res 912:33–46
6. Mausset-Bonnefont AL, Hirbec H, Bonnefont X, Privat A, Vignon J, de Seze R (2004) Acute exposure to GSM 900-MHz

electromagnetic fields induces glial reactivity and biochemical modifications in the rat brain. Neurobiology 17:445–454
7. Okatan DÖ, Okatan AE, Hancı H, Demir S, Yaman SÖ, Çolakoğlu S, Odacı E (2018) Effects of 900-MHz electromagnetic fields exposure throughout middle/late adolescence on the kidney morphology and biochemistry of the female rat. Toxicol Ind Health 34(10): 693–702
8. Odaci E, Bas O, Kaplan S (2008) Effects of prenatal exposure to a 900 MHz electromagnetic field on the dentate gyrus of rats: a stereological and histopathological study. Brain Res 1238:224–229
9. İkinci A, Odacı E, Yıldırım M, Kaya H, Akça M, Hancı H, Aslan A, Sönmezll OF, Baş O (2013) The effects of prenatal exposure to a 900-megahertz electromagnetic field on hippocampus morphology and learning behavior in rat pups. NeuroQuantology 11(4):582–590
10. Narayanan SN, Kumar RS, Potu BK, Nayak S, Bhat PG, Mailankot M (2009) Effect of radio-frequency electromagnetic radiations (RF-EMR) on passive avoidance behaviour and hippocampal morphology in Wistar rats. Ups J Med Sci 115(2):91–96
11. Eyre MD, Richter-Levin G, Avital A, Stewart MG (2003) Morphological changes in hippocampal dentate gyrus synapses following spatial learning in rats are transient. Eur J Neurosci 17: 1973–1980
12. Tong J, Chen S, Liu XM, Hao DM (2013) Effect of electromagnetic radiation on discharge activity of neurons in the hippocampus CA1 in rats. Zhongguo Ying Yong Sheng Li Xue Za Zhi 29(5):423–427
13. Wang H, Peng R, Zhou H, Wang S, Gao Y, Wang L, Yong Z, Zuo H, Zhao L, Dong J, Xu X, Su Z (2013) Impairment of long-term potentiation induction is essential for the disruption of spatial memory after microwave exposure. Int J Radiat Biol 89(12):1100–1107
14. Adebayo EA, Adeeyo AO, Ogundiran MA, Olabisi O (2018) Biophysical effects of radiofrequency electromagnetic radiation (RF-EMR) on blood parameters, spermatozoa, liver, kidney and heart of albino rats. J King Saud Univ - Sci 31(4):813–821
15. Bas O, Odaci E, Kaplan S, Acer N, Ucok K, Colakoglu S (2009) 900 MHz electromagnetic field exposure affects qualitative and quantitative features of hippocampal pyramidal cells in the adult female rat. Brain Res 1265:178–185
16. Kishore GK, Venkateshu KV, Sridevi NS (2019) Effect of 1800-2100 MHz electromagnetic radiation on learning-memory and hippocampal morphology in Swiss albino mice. J Clin Diagn Res 13(2):14–17
17. Lu Y, Xu S, He M, Chen C, Zhang L, Liu C, Chu F, Yu Z, Zhou Z, Zhong M (2012) Glucose administration attenuates spatial memory deficits induced by chronic low-power-density microwave exposure. Physiol Behav 106(5):631–637
18. Razavinasab M, Moazzami K, Shabani M (2014) Maternal mobile phone exposure alters intrinsic electrophysiological properties of CA1 pyramidal neurons in rat offspring. Toxicol Ind Health 32(6):968–979
19. Pfefferbaum A, Mathalon DH, Sullivan EV, Rawles JM, Zipursky RB, Lim KO (1994) A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. Arch Neurol 51(9):874–887
20. Paulraj R, Behari J (2006) Single strand DNA breaks in rat brain cells exposed to microwave radiation. Mutat Res 596:76–80
21. Kesari KK, Behari J (2009) Fifty-gigahertz microwave exposure effect of radiations on rat brain. Appl Biochem Biotechnol 158: 126–139
22. Kesari KK, Behari J (2010) Effect of microwave at 2.45 GHz radiations on reproductive system of male rats. Toxicol Environ Chem 92:1135–1147
23. Pandey UB, Nichols CD (2011) Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery. Pharmacol Rev 63(2):411–436
24. Mañas P, Mackey BM (2004) Morphological and physiological changes induced by high hydrostatic pressure in exponential- and stationary-phase cells of Escherichia coli: relationship with cell death. Appl Environ Microbiol 70(3):1545–1554
25. Gonzalez RC, Woods RE (2001) Digital image processing. Prentice-Hall, NJ
26. Gonzalez RC, Woods RE, Eddins SL (2004) Digital image processing using MATLAB. Pearson Education. ISBN 978-81-7758-898-9
27. Khan SS, Ahmad A (2004) Cluster centre initialization algorithm for K-means clustering. Pattern Recogn Lett 25:1293–1302
28. Yi B, Qiao H, Yang F, Xu C (2010) An improved initialization center algorithm for K-means clustering. IEEE
29. Tan F et al (2008) A genetic algorithm-base3d method for feature subset selection. Soft Comput 12:111–120
30. Liu H, Setiono R (1995) Chi2: feature selection and discretization of numeric attributes. Proc 7th IEEE Int Conf Tools with Artif Intell 388–391
31. Xue M, Zhang W, Browne N, Yao X (2016) A survey on evolutionary computation approaches to feature selection. IEEE Trans Evol Comput 20(4):606–626
32. Mohammadi M, Hofman W, Tan Y-H (2018) A comparative study of ontology matching systems via inferential statistics. IEEE Trans Knowl Data Eng 31:615–628. https://doi.org/10.1109/TKDE.2018.2842019

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Ritesh Maurya is currently a PhD candidate in Department of Computer Science & Engineering, Centre for Advanced Studies, AKTU, India. He received M. Tech degree from the Indian Institute of Information Technology & Managment, Gwalior.



Ms. Neha Singh is holding M. Tech in Environmental Engineering and is pursuing a PhD in Environmental Sciences from Amity University, Noida. She has expertise in EMF radiation, biological interactions, and genetics.



Prof. Tanu Jindal is PhD from Delhi University. Currently, she is serving as Gp. Pro-Vice-Chancellor (R&D), Amity University, Noida. She has 6 patents, three books, and over 50 publications in refereed journals to her credit.



Prof. M. K. Dutta is currently working as a Professor in Centre for Advanced Studies, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India, and also is the Dean Post-graduate Studies and Research of the university. His research interests include machine learning applications mainly in biology and medicine.



Prof. Vinay K Pathak did his B. Tech. in Computer Science from Harcourt Butler Technological Institute, Kanpur in 1991; M. Tech. from IIT Kharagpur in 1998; and PhD in Computer Science from U.P. Technical University, Lucknow under the joint guidance of IIT Kanpur, November 2004. He is currently working as Vice-Chancellor, Dr. A.P.J. Abdul Kalam Technical University, U.P., Lucknow, India. Research interests of Prof. Pathak include computational geometry and image processing.