The Visual Computer (2023) 39:1815–1827
https://doi.org/10.1007/s00371-022-02447-9

ORIGINAL ARTICLE

Gender identification of *Drosophila melanogaster* based on morphological analysis of microscopic images

Leonid M. Mestetskiy[1] · D. S. Guru[3] · J. V. Bibal Benifa[2] · H. S. Nagendraswamy[3] · Channabasava Chola[2]

Accepted: 20 February 2022 / Published online: 19 March 2022

Abstract

Drosophila melanogaster (*D. melanogaster*) is an imperative genomic model organism that is employed widely in healthcare and biological research works. Roughly 61% of recognized human genes have a perceptible similarity with the genetic code of *D. melanogaster* flies, besides 50% of its protein structures have mammalian equivalents. In recent times, numerous studies have been done in *D. melanogaster* to investigate the functions of particular genes that are available in its central nervous system, including the major organs like the heart, liver and kidney. The findings of these research works through *D. melanogaster* are utilized as a key mechanism to explore human interrelated diseases. However, it is essential to recognize the male and female Drosophila flies for the better understanding of human disease related studies, and it is a tricky job. This paper describes a unique programmed system to categorize the gender of *D. melanogaster* from the ventral view portraits captured through microscope. The proposed method includes image segmentation of the body of *D. melanogaster* in the form of a binary image and the construction of a continuous morphological model based on its skeleton. An analysis of the skeleton makes it possible to assess the sharpness of the caudal end of the *D. melanogaster* abdomen through a detailed assessment of the curvature. Based on this assessment, a *Drosophila melanogaster* Gender (DMG) classifier is constructed for the gender determination of *D. melanogaster* flies. The accuracy of the DMG classifier is about 98% in proportion to the existing state-of-the-art shape-based classifiers with optimal computing time.

Keywords *Drosophila melanogaster* · Body shape · Morphological model · Skeleton · Gender recognition

1 Introduction

In recent days, the accuracy of image processing algorithms have been improved significantly toward image classification, object detection and object tracking from the images and videos. The recent developments of image processing methods with Machine Learning (ML) have played a vital role in the area of biomedical engineering research. The biomedical research includes cell classification, cell counting and live cell imaging from microscopic image analysis. Microscopic image classification is one of the most useful techniques to characterize the outcomes of biology-based research activities.

The gender identification of *D. melanogaster* by means of microscopic images is a complex problem, and it is essential for various genetic studies. However, a subject expert only could characterize this problem based on the microscopic visual analysis that is guided by several predefined observation features. Hence, the development of a classification algorithm is mandatory to automate the recognition process during the mass analysis of a large number of images. The gender of *D. melanogaster* flies can be differentiated visually through the size (dimensions) and shape of the body. The dimensions of the body are related with the genotype and phenotype of the flies, and conversely, it influences the gender classification process (Handa, 2014) [1]. Further, in the course of gender classification of *D. melanogaster*, the size polymorphism is one of the exciting tasks. It should be noted that the female flies are slightly larger than males; however, the size of the fly depends largely on the nutritional conditions, especially at the larval stage.

✉ J. V. Bibal Benifa
benifa.john@gmail.com

1 Department of Computational Mathematics and Cybernetics, Moscow State University, Moscow, Russia

2 Department of Computer Science and Engineering, Indian Institute of Information Technology, Kottayam, India

3 Department of Studies in Computer Science, University of Mysore, Mysore, Karnataka, India

Earlier, various techniques have been projected to classify the male and female *D. melanogaster* flies according to the wing texture of both genders by considering the features such as local and transformed local binary patterns. In addition, sign and magnitude features along with Support Vector Machine (SVM) and Random Forest (RF) algorithms are also considered for the classification process [2]. They have achieved an average accuracy of about 84% for the gender classification of *D. melanogaster*. Later, Neto et al. (2017) suggested a gender classification strategy with respect to the fractal dimension and wing texture of the flies. The concept of Stationary Wavelet Transformation (SWT) is used to extract the core features such as canny filter and fractal dimension from the microscopic images. Here, SVM and RF classifiers were implemented to accomplish the classification task successfully [3].

Drosophila melanogaster has pronounced gender dimorphism such that the females and males morphologically differ from each other in a number of features [1]. Primarily, the female abdomen is slightly rounded with a pointed end (Fig. 1a), and in the male, it appears like a cylindrical blunt end as highlighted in Fig. 1b. These features are clearly distinguishable from microscopic images during the visual analysis. Other morphological features are associated with differences in pigmentation, the number of chitinous bristles of the breast (tergites) and chitinous plates on the ventral side (sternites).

The visual symptoms of gender differences should be considered when developing an algorithm for classifying the gender of *D. melanogaster* from the microscopic images. However, from the microscopic images the accurate determination of true size and elements of *D. melanogaster* pigmentation is a challenging process because of the position of the flies in the frame, the shooting range and lighting conditions. Accurate identification of tergites and sternites from the microscopic images is also a complex process due to insufficient resolution and occlusions. In the view of stated reasons, the most reliable feature to be considered for gender identification is the shape of the abdomen. In this article, a model for classifying and identifying the gender of *D. melanogaster* based on the abdomen shape analysis is presented. The highlights of the proposed research are summarized as follows:

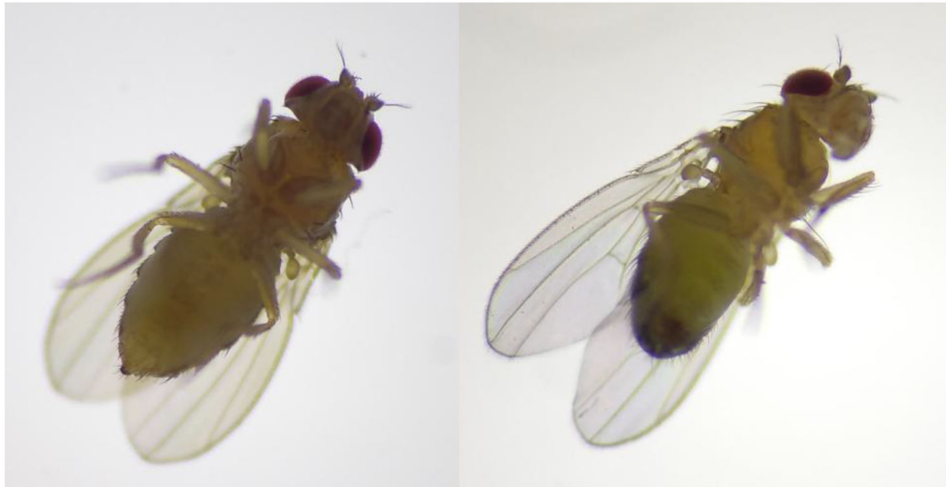
- The public or standard dataset with a ventral view of *D. melanogaster* flies is not available as on date in any repositories. Hence, a novel dataset with microscopic images of *D. melanogaster* flies is constructed under various lighting conditions.
- A method is proposed for solving the gender classification problem based on the abdomen shape analysis of the *D. melanogaster*.
- It uses a continuous morphological model that includes the geometrical approximation of contour of the body of a *D. melanogaster* by a polygonal figure and the construction of the middle axis (skeleton).
- To carry out the image segmentation process, the color images available in the dataset are transformed into binary, and a threshold limit is fixed. Then, the construction of a skeleton graph with polygon figures is done with the help of computational geometry that adapts to a continuous morphological model.
- Finally, the classification process is done based on the set of incident inscribed circles by calculating the abdominal border line contours.

The subsequent sections of the article are structured as follows: Sect. 2 describes the contemporary literature review with respect to the shape analysis of *D. melanogaster*. In Sect. 3, the proposed method is presented followed by the creation of a dataset, experimental results and discussions in Sect. 4. The key inferences of the study and the conclusions of the work carried out are summarized in Sect. 5.

2 Related works

Drosophila melanogaster fly is deemed to be one of the impressive model organisms for studying the principal sources of diseases related to human beings [4]. It is widely adopted for medicinal investigations owing to the reality that the fruit fly's inner physiological elements, and their functions are analogous to the ones in vertebrates, as well as human beings. Minor dissimilarities present in the midst of humans and *D. melanogaster* with reference to their gross morphological and cellular features, though various physiological, biological and neurological characteristics are preserved in-between the two organisms. Further, it is also essential to note that approximately 75% of human disease-oriented genes are supposed to possess functional resemblance with *D. melanogaster* flies [5]. Neurological disorders, cancer, cardiovascular disease, metabolic and Lysosomal storage disorders, along with the genes necessary for the proper functioning of the optical, auditory and antibodies are also identical to the protein sequences of *D. melanogaster* [6]. Recently, quite many assays have been advanced on *D. melanogaster,* and they were specifically employed for the research on human-related syndromes. The studies comprise of central nervous system-based evaluations such as listening, absorbing, reminiscence and periodic movements have been performed in *D. melanogaster*. The outcomes of evaluation are utilized to understand the peripheral nervous system dysfunction, with neurodegeneration, dementia, seizure disorder, apoplexy, traumatic brain injury

Fig. 1 Expanded microscopic images of *D. melanogaster*



(a). Female (b). Male

and tumors [7]. *Drosophila melanogaster* is utilized as a prototypical organism to examine Alzheimer's disorder, and it is also associated with deterioration in cognitive functions, hikikomori, dementia and long/short-term memory problems [8]. Although, the fruit flies have not been used previously in asthma-related studies, *D. melanogaster* might be extremely helpful in relating genetic developments to biological functions [9].

The gender differentiation of *D. melanogasters* can be carried out based on the outline contours or anatomical structure of the body as presented in Fig. 1. Further, the dimensions of the bodily skeleton are fully related to the genotype and phenotype of the fruit flies that ultimately influences the gender identification process. Handa et al. (2014) indicated that the dimension of flies show the functions in contrast for the two genders due to the existence of polymorphism (heterogeneity) in *D. melanogaster* [1]. The size heterogeneity is one of the challenges, and it affects the gender classification process. In the last decade, various methods have been recommended to differentiate the gender of *D. melanogaster* according to the wing configurations of the fly [2]. Here, wing texture characteristics (sign and magnitude) of both fruit flies are reproduced through local and modified local binary patterns along with SVM and RF classification algorithms. Ahmad et al. (2014) attained 84% mean accuracy for the gender-specific classification via wing texture feature elicitation process. Texture is a feature employed to split the images into specific zones of interest for further classification of those regions from an image insert [2]. Neto et al. (2017) suggested a unique gender classification model according to the geometric features of wings using SWT and canny filters. In addition, they have implemented SVM and RF algorithms for the classification assignments [3]. Later, Priyanka and Sudhakar (2018) employed the Histogram of Oriented Gradients (HOG) for the retrieval of shape features and K-Nearest Neighbor (KNN) classifiers to group the categories depending on the spatial affiliations [10].

Shape is a significant visual characteristic of an image that is essential for the human visual system to classify and recognize the objects [11]. Shape depends on the boundary profiles of objects and its interior characteristics [12–15]. Shape analysis aims to determine a distance metric to measure the degree of dissimilarity among the different shapes. Shape-based features are being used in the field of biomedical engineering, text analysis, microscopic image analysis, hyperspectral image analysis and computer vision for object detection, classification and retrieval purposes [16–20]. Alsmadi (2020) proposed a Content-Based Image Retrieval (CBIR) system for coral dataset with 1000 images. The images are retrieved with the help of shape and texture features using a Metaheuristic-based algorithm [12]. A well-known Convolutional Neural Network (CNN) architecture namely Visual Geometry Group (VGG) is used for extracting the shape features and to improve the classification performance [21]. Similarly, classification of lung nodules is efficiently done by an ensemble classifier known as AdaBoosted Back Propagation Neural Network that extracts deep texture and shape features [14].

Later, significant work has been reported for shape recognition and contour extraction in the Moving Picture Experts Group-7 (MPEG-7) dataset. Here, morphological methods are used for feature extraction, and SVM is used for classification that resulted in a good recognition rate about 98.21% [18]. Subsequently, Huixian (2020) identified the plant species by extracting the shape and texture features from 50 different kinds of leaf databases. Here, KNN, SVM

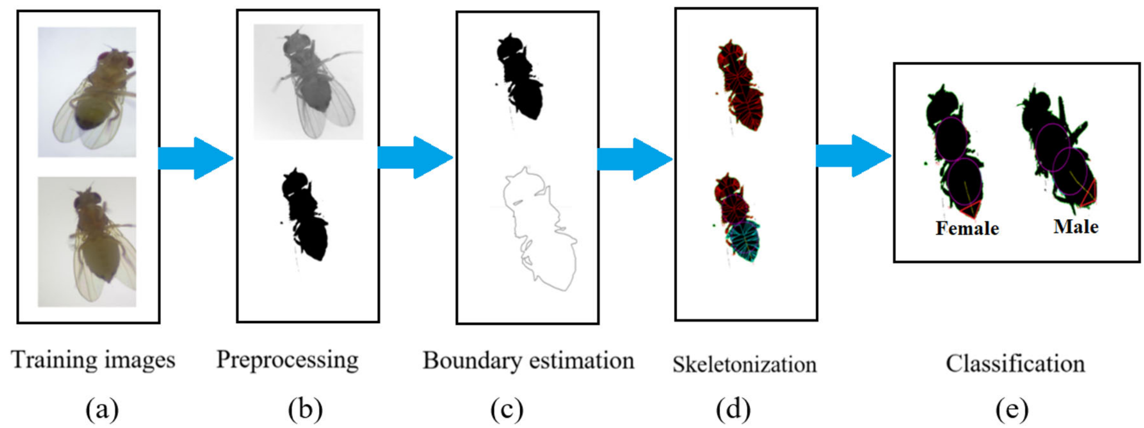


Fig. 2 High-level flow diagram of DMG classifier

and Artificial Neural Network (ANN) are used for leaf image recognition purposes, and ANN achieves an average recognition rate of about 92.47% in 122 ms [19]. Biomedical engineering has received more attention in recent days because many Deep Learning (DL) techniques are adopted in various domains such as medical image segmentation, classification and detection. Several well-known DL architectures such as ResNet [22], SqueezeNet [23], InceptionV3 [24], AlexNet [25], DarkNet [26], VGG [27] and Unet [28] are being used as transfer and learning techniques. Mammogram detection, segmentation and classification are done by You Look Only Once (YOLO) which uses CNN as base architecture for the feature extraction and the detection of Region Of Interest (ROI). Here, AlexNet-based deep CNN is used for classification of images into benign or malignant [29]. Further, few other works have reported Adult mosquitoes classification based on their morphology with the help of pre-trained Neural Networks (NN) such as AlexNet, GoogleNet and LeNet [20]. Here, the classification of species and mosquito gender identification are done with the help of DL techniques [20].

Rebelo et al. (2020) presented an extensive survey on automatic image-based species classification of winged insects namely fruit flies, butterflies and mosquitoes [30]. Majority of the reviewed works have employed image processing, computer vision and ML techniques. The authors also reported the scarcity of public datasets, and due to the computational challenges, only sparse amounts of research works are done in such domains. Beetle and wasp gender and species identification are also done with the help of RF, SVM, Logistic Regression (LR), multilayer perceptron and AdaBoost methods [31]. However, LR exhibited a better performance as compared to other ML techniques in terms of Beetle gender and species identification problems. Li and Xiong (2017) proposed a method for automatic butterfly species identification based on Histograms of Multi-Scale Curvature (HoMSC) and Gray-Level Co-occurrence Matrix of Image Blocks (GLCMoIB) [32]. Here, HoMSC and GLCMoIB are used to explain the butterfly wing's shape and texture characteristics, respectively. Li and Xiong (2017) utilized totally 750 images of 50 various butterfly classes that are utilized for the classification studies. In addition to this, a weight-based KNN learning algorithm is also employed, and the obtained accuracy rate is about 98%.

From the literature, it is observed that only sparse research works have been done till date in the specialization of *D. melanogaster* gender classification by computer-aided learning techniques. Further, Channabasava et al. (2022) have presented the *D. melanogaster* gender classification using ventral image datasets and ML techniques [33]. However, the employment of ML techniques are not cost effective for real-time small scale applications as it requires more computing power and memory. Hence, in the presented work, a light-weight model (shape-based classifier with the executable file size of $\leq$ 2 kB) is developed using morphological analysis of *D. melanogaster* ventral view images. This shape-based classifier excels in the classification process and gender determination of *D. melanogaster* from the microscopic image sources, and various classification techniques are also compared with a validation perspective.

3 Proposed methodology

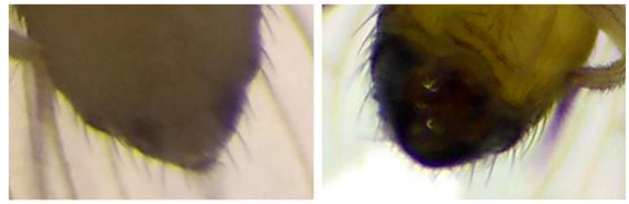
A method for solving the gender identification problems in *D. melanogaster* is proposed based on the abdomen shape analysis of the *D. melanogaster* using a continuous morphological model. The high-level work flow of the proposed *D. melanogaster* gender (DMG) identification model is displayed in Fig. 2.

The proposed approach includes the following steps:

- *D. melanogaster body segmentation from microscopic images*

(a) Female (b) Male

Fig. 3 Border curvature of the caudal part of *D. melanogaster* abdomen



(a) Female (b) Male

Fig. 4 Enlarged image of the end portion of *D. melanogaster* abdomen

The original color image shown in Fig. 2a is converted into a binary image, in which the body of the fly is highlighted in black color (Fig. 2b).

- *Construction of a continuous morphological model from the binary image*
  A polygonal figure is constructed that approximates the body of a *D. melanogaster* in a binary image with high accuracy. For the obtained polygonal Fig. 2c, a skeleton is constructed using the methods of computational geometry.
- *Determination of D. melanogaster body parts and the abdominal centerline*
  The Fly body parts and abdominal centerline are determined by analyzing the skeletal graph and the set of incident inscribed circles as shown in Figs. 2d and 6e, respectively.
- *Determination of the border of abdomen region*
  The boundary is the envelope of a family of circles whose centers lie on the centerline of the abdomen region.
- *Assessment of the pointedness of the abdomen caudal end of D. melanogaster*
  Here, the sharpness is described by the assessment of the curvature of the abdominal border line. Comparison of the pointedness indicator with the threshold is the essence of the decisive rule for the gender classification of *D. melanogaster*.

The proposed method is implemented and experimentally tested based on the indigenously developed dataset of 400 images. The proportion of accurate determination of the gender of *D. melanogaster* was achieved about 98%.

3.1 Body shape of *D. melanogaster*

The male and female fruit flies can be differentiated by recognizing the boundary line of the caudal part of the abdomen as highlighted in Fig. 3. In the male, this line has a slightly smooth curvature, and in the female, the boundary line "breaks" at the tail vertex point.

Several complexities exist in the automated identification of male and female flies through ML techniques. Firstly, the abdomen portion of the image does not contain a clearly defined border, and it is clearly visible as the image is scaled up (Fig. 4). Secondly, the abdomen is covered with chitinous setae that look like hairs. Both of these factors significantly distort the image at the border line of the abdomen. Thirdly, the position, orientation of the front view and the camera angle are widely varying, and the determination of abdomen curvature would be a tricky problem. Ultimately, these factors significantly complicate the exact segmentation and classification of the abdomen shape of *D. melanogaster* from the input dataset.

3.2 Development and analysis of a continuous morphological model

The solution is to construct an exact border line, evaluate its curvature, and based on the borderline curves, it should be determined whether the tail section is pointed or rounded one. Hence, the solution is primarily focused on the construction of a continuous morphological image model [34]. A continuous morphological model allows constructing a curve that describes the shape of the end part of the *D. melanogaster* abdomen. It also evaluates whether the edge is rounded or pointed one with reasonable accuracy. This process includes the following steps as stated in the above Sect. 3.

3.2.1 *Drosophila melanogaster* image segmentation

The purpose of segmentation is to highlight the silhouette of *D. melanogaster* in the microscopic images. The body of the fly must be separated from the background as well as from the wings. The choice of segmentation method substantially depends on the format, resolution and quality of the source images. In the proposed experiments, segmentation is performed by selecting a threshold based on the construction and analysis of the brightness histogram as shown in Fig. 5. Two local maxima are well distinguished on the histogram as highlighted in Fig. 5b. The left maximum corresponds to the

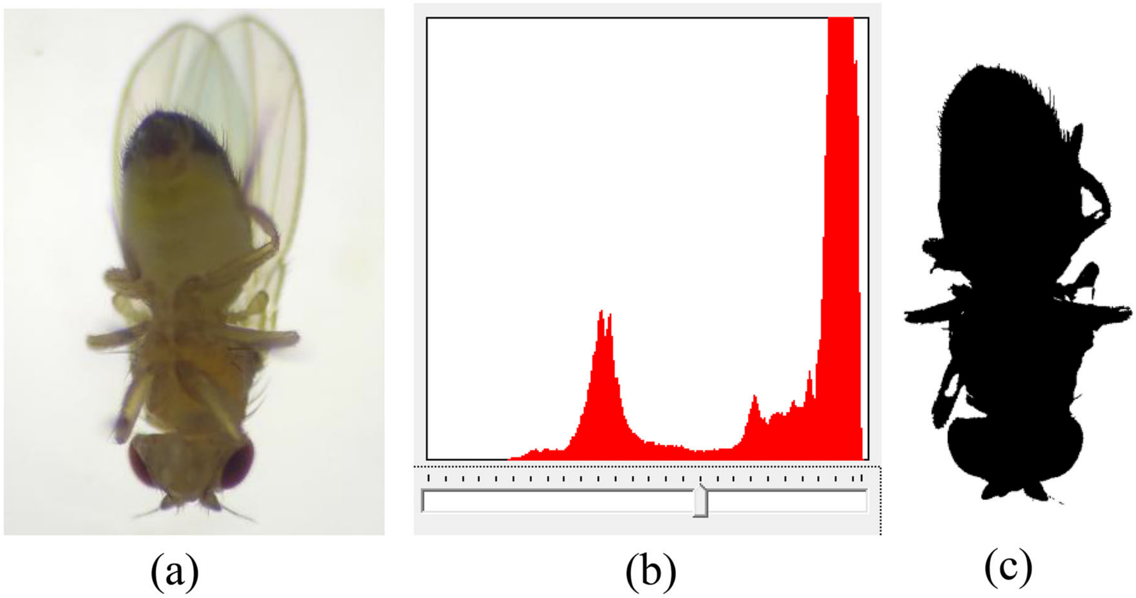


Fig. 5 *Drosophila melanogaster* body segmentation: **a** source image, **b** blue histogram of brightness, **c** binary image

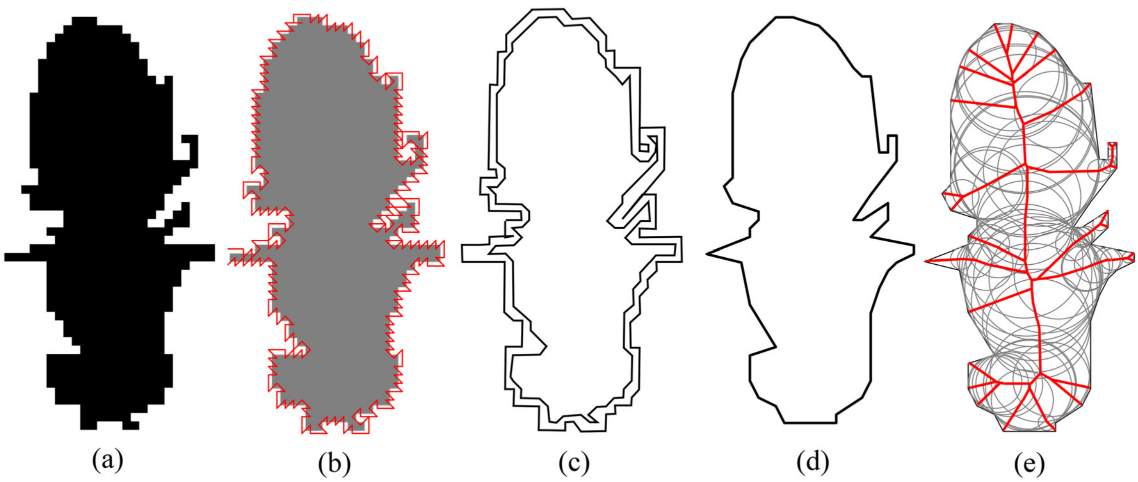


Fig. 6 Construction of a CMM of the silhouette of a *D. melanogaster*: **a** a binary image, **b** a trace of a border tracking, **c** a boundary corridor, **d** a polygonal figure, **e** inscribed circles and skeleton

dark colors of the *D. melanogaster* body, and the right shows the light background colors. There is a well-defined local minimum between them, which is chosen as a threshold for binarization. The resulting binary image highlights the body of the fly by cutting off the background and wing sections. Result of the segmentation process is a binary image in which the silhouette of a body of a fly without wings is depicted in black on a white background as shown in Fig. 5c.

3.2.2 Building a continuous morphological model

A Continuous Morphological Model (CMM) of a binary image (Fig. 6a) of a *D. melanogaster* silhouette consists of a polygonal Figure as represented in Fig. 6d. It is attained by approximating the silhouette border, a set of circles inscribed in the skeleton as displayed in Fig. 6e. A polygonal shape is simply a polygon with multilateral holes, and it will be formed, if the resulting silhouette is not simply connected. The skeleton is a group of points whose centers imprinted in a polygonal image of empty circles. An empty circle is called inscribed, if it is not contained entirely inside another empty circle. The construction of an approximating polygonal figure and its skeleton is carried out using the methods described by Mestetsky, (2009) [34].

As an example case, the operation of building CMM is illustrated in Fig. 5, and the input of the model is a bitmap binary image (Fig. 6a). As the proposed model is competent enough to handle low as well as high resolution images, a

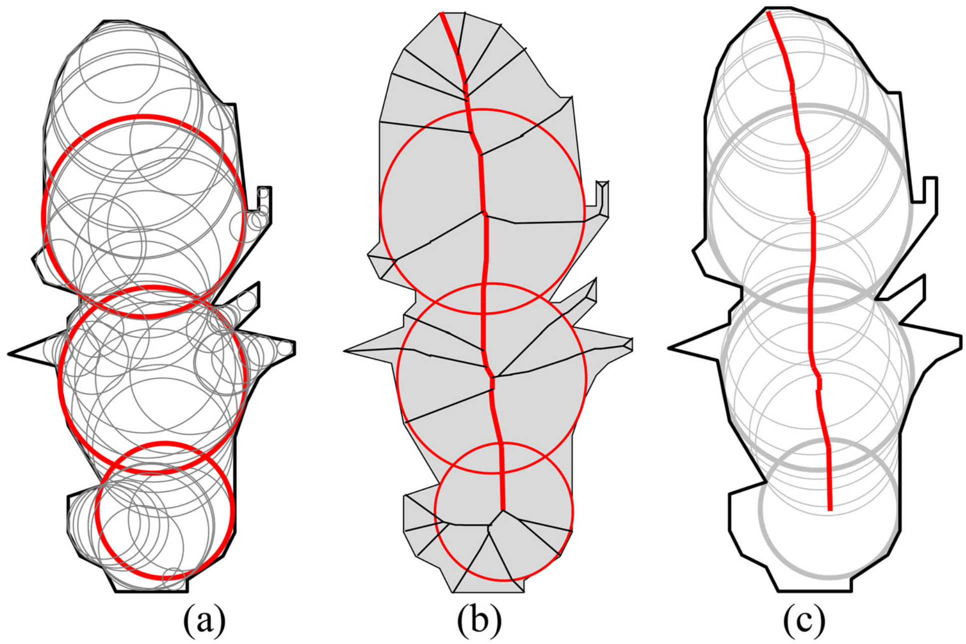


Fig. 7 *Drosophila melanogaster* body segmentation: **a**—selection of the head, chest and abdomen, **b**—selection of the silhouette axis, **c**—selection of circles with centers on the axis

low-resolution raster image of the *D. melanogaster* body was selected, and the border is traced for visualization (Fig. 6b). Subsequently, the boundary corridor is estimated as represented in Fig. 6c. The output of the algorithm describes the boundary contour of an object in the form of a polygon as displayed in Fig. 6d, and the medial illustration of the polygonal view is displayed in Fig. 6e. The step-by-step method of construction used for approximating the polygonal figure including principles are given below:

- As the geometry of *D. melanogaster* body resembles a digital image, a subset of the integer lattice $\mathbb{Z}^2$ is considered on the Euclidean plane that is enclosed by a rectangle. Hence, a binary image is developed, and its pixels are illustrated as distinct black & white grid points. The points within the object are represented as black, and the points in relation to the background are displayed in white color.
- Next, a trilateral neighborhood pattern is introduced on the lattice $\mathbb{Z}^2$. The neighborhood chart is prepared by triangulation process in which the square elements of an integer lattice are splitted diagonally. The set of all wedge-shaped cells accompanied by multi-colored apogees represents an annular strip known as the boundary corridor as highlighted in Fig. 6c.
- In all closed loops located inside the boundary corridor, a line of least dimension is selected. The shortest line has the shape of a simple polygon, and the convex vertices of the polygon are represented as black pixels, while the concave vertices are in white pixels. The polygon separating the dark and snowy pixels of the boundary corridor is known as the untying polygon of the least perimeter (Fig. 6d).

To obtain the boundary corridor precisely, the algorithms for tracing the boundary of a binary image are used [34]. The task of constructing the separating polygon of the minimum perimeter is fulfilled by the shortest ring route inside the boundary corridor. It is similar to the familiar problem of creating the defined geodetic path in the polygon (Sklansky, 1972) [35]. In the present research, the resulting polygon is being formed from the triangulated polygons by an effective algorithm through the construction of a geodesic path (Toussaint, 1989) [36]. Subsequently, by triangular lattice a solution is arrived easily using the “thread pulling” method [34]. The medial representation of a polygonal figure is computed using the effective algorithms of computational geometry that are already developed by the authors of this work [34, 37].

3.2.3 Structural analysis of the body of *D. melanogaster*

Drosophila melanogaster body parts—head, chest and abdomen are distinguished based on the analysis of the size of the inscribed circles. The skeleton can be considered as a geometric graph where each point has a weight equivalent to the radius of the marked circles centered at that point. This weight is certainly known as the radial function of the skeleton, and it has three local maxima in the silhouette of *D. melanogaster*: (i) head, (ii) chest and (iii) abdomen. The parts of the *D. melanogaster* body are localized with the help of the inscribed circles corresponding to the maxima as presented in Fig. 7a.

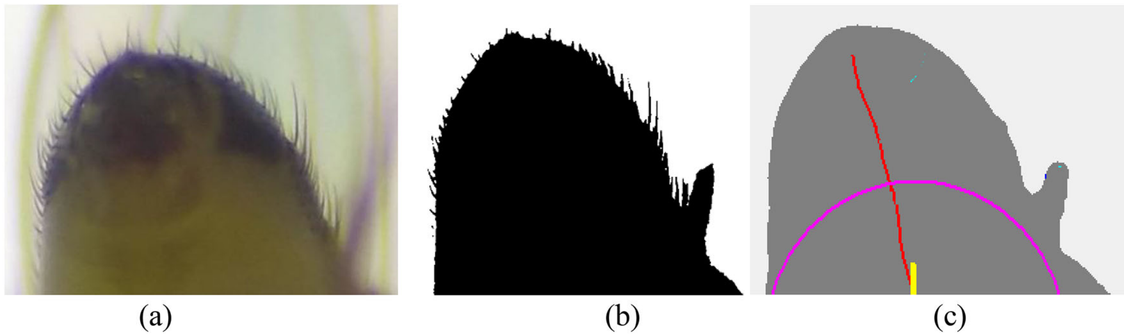


Fig. 8 Determination of the abdominal boundary line: **a** a fragment of the original image, **b** a binary image, **c** the union of a family of inscribed circles with centers on the silhouette axis

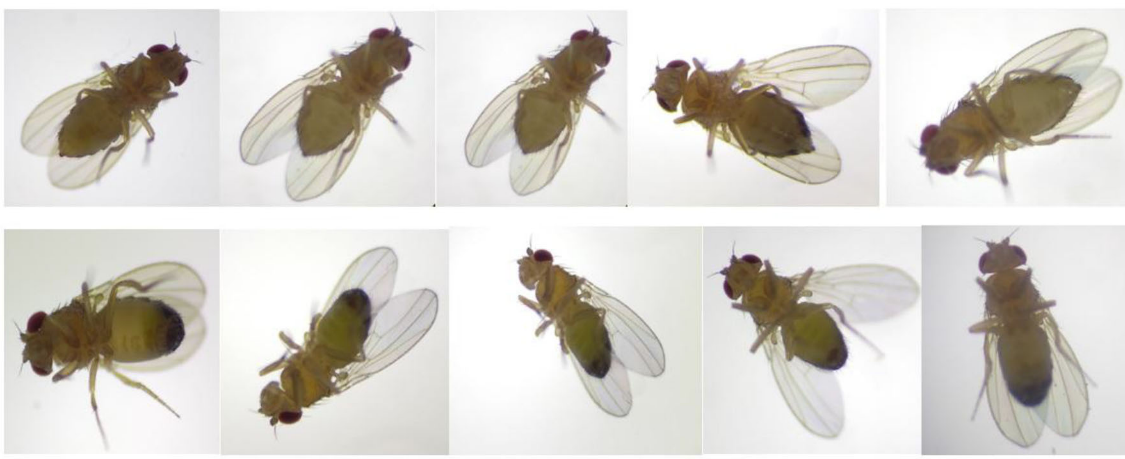


Fig. 9 Examples photographs of *D. melanogaster* under a microscope

The head portion corresponds to the smallest among the three circles as shown in Fig. 7a. This feature makes it possible to determine the position of the head, chest and abdomen regions as well as to determine the orientation of the *D. melanogaster* body. The vector direction extends from the center of the chest to center of the abdomen. Using this vector, the skeleton branch is easily distinguished, which is the axial line of the abdomen as displayed in Fig. 7b.

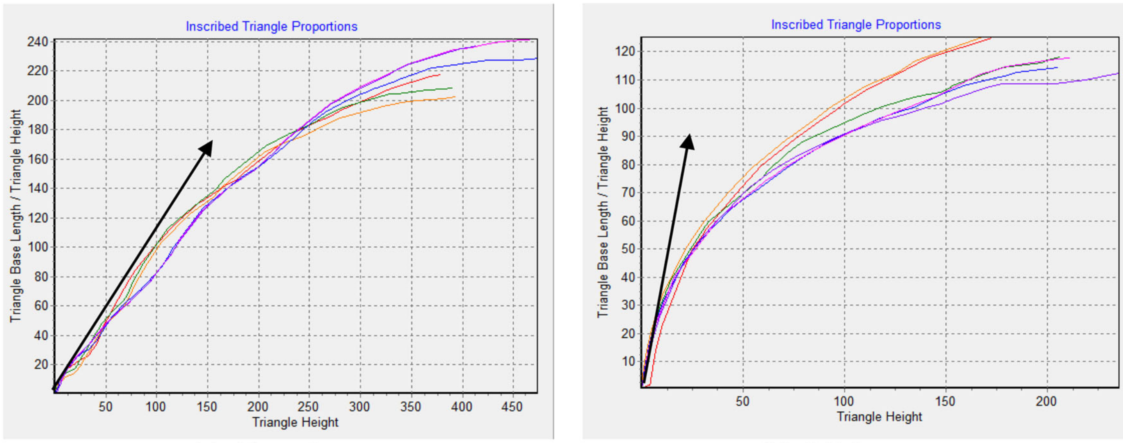
3.2.4 Determination of the abdominal boundary line

The abdominal border line is calculated on the basis of smoothing the edges of the image. However, qualitatively it is not possible by means of raster morphological operations such as dilatation and erosion. In the present work, this problem is solved by extracting all the inscribed circles having centers on the axis of the abdomen and subsequently by calculating the envelope for this family of circles (Fig. 7c). Figure 7 illustrates the idea of a smoothing method using a model example of a low-resolution image.

A real example of constructing the abdominal border line using the envelope of a family of inscribed circles whose centers on the silhouette axis is shown in Fig. 8. For the initial input image (Fig. 8a), a binary image was obtained as a result of the segmentation process (Fig. 8b). Using a binary image, a medial representation is created with axis selection, and a union of a family of inscribed circles with centers on the axis is constructed (Fig. 8c). The envelope of the family of circles is considered as the boundary of the caudal part of the abdomen. This boundary line is then used to identify the *D. melanogaster* gender.

4 Experimental setup and results

To carry out the experimental analysis, the present research work is executed in the MATLAB 2019 (a) environment that is deployed in a computer consisting of an *i7* processor with 32 GB Random Access Memory (RAM) capacity. As the standard *D. melanogaster* image dataset (ventral view) is unavailable openly, the *D. melanogaster* ventral views were captured at the National *Drosophila* stock centre, University of Mysore, India. Totally, 400 microscopic images with two different classes (male—200 and female—200 images) are



(a) Female (b) Male

Fig. 10 Determination of the pointedness of the abdominal boundary line

compiled as a dataset with a digital compound Microscope facility. The dataset contains samples of normal and non-mutated (wild types) strains of *D. melanogaster.* These *D. melanogaster* images are captured using digital compound Microscopes at various lighting circumstances which cause several difficulties such as dissimilarity in viewpoint, brightness and background reflections. The essential depictions of active fruit flies are taken by anesthetizing and holding it on the platform mounted glass slide of compound Microscope at an ideal standpoint. The model images of both *D. melanogaster* gender flies are presented in Fig. 9 where the top and bottom rows consist of female and male flies, respectively.

In this section, the results corresponding to the *D. melanogaster* gender classification by different ML and DL classifiers are presented. For the experimental analysis, an exclusive dataset is developed, and only the ventral observations of *D. melanogaster* images are used for initializing the solution. After preprocessing, the resolution of each image in the dataset is maintained as 227 $\times$ 227. The only images included for preliminary processing are subjected to the feature extraction process, and the educed features are assigned to the supervised learning methods. Supervised learning techniques require training sample cases with designated data to perform the classification functions, and the competence of the suggested method is ensured by various accuracy metrics.

The various methods such as SVM, CNN and BPNN are matched with the proposed technique in respect of accuracy and time consumption for classifying the images. To validate the suggested model with the existing models, the dataset is divided into 70:30 (70% for training and 30% testing) for the classification process. Each model is iterated 10 times in order to ensure the reliability of results.

4.1 *Drosophila melanogaster* abdomen borderline curvature assessment

Based on the obtained boundary line, abdominal width diagrams are plotted relative to the axis (Fig. 10). The distance from the endpoint of the abdomen to the point on the axis of the abdomen is plotted along the abscissa axis. Then, the radius of the marked circle centered at this point is positioned along the ordinate axis, and the resulting curve depicts the border of the abdomen. The sharpness of the edge can be estimated based on the angle of tangent to this curve at the origin. Further, the angle is estimated numerically by approximating the resulting curve with a 3-order polynomial expression.

As a result of the above actions, the angular coefficient of the tangent to the abscissa axis is obtained for the original *D. melanogaster* image. Figure 10a and b shows the typical lines of the envelope corresponding to the abdomen curvature of females and males, respectively. In this work, based on the largest inscribed circle the pointedness of the fly is computed by the help of triangle estimation. Based on the triangle estimation value, the angular coefficient is calculated. Figure 10 shows that the angular coefficient of the tangent of females is significantly less than that of males. This property clearly reflects the differences in the abdominal shape of *D. melanogaster*. To build a decision rule, a threshold value of the angular coefficient is selected. It should be noted that the numerical value of this coefficient is selected on the basis of computational experiments. The threshold value was selected based on experiments with a database of 200 females and 200 males. In this work, the resulting threshold value of the angular coefficient is equal to 1.7.

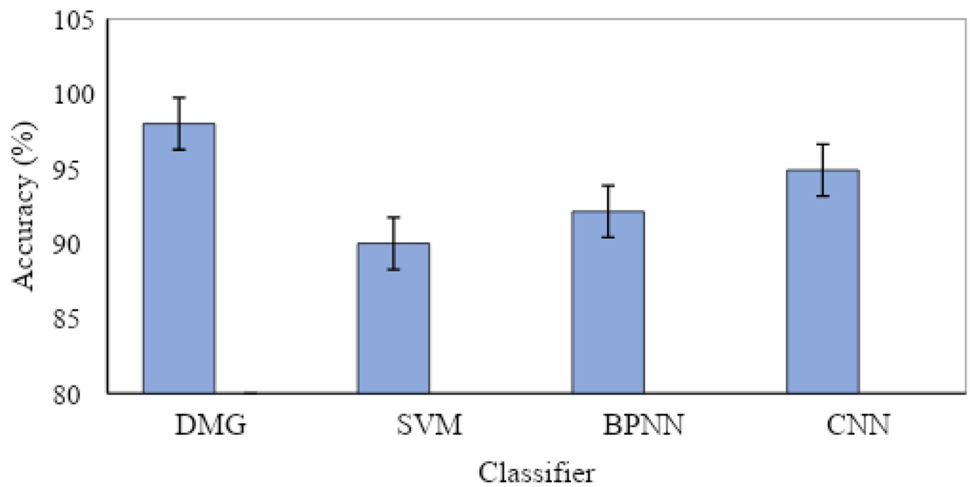


Fig. 11 Accuracy analysis of different classifiers with error bars (standard deviation)

4.2 Comparison with the state-of-the-art learning techniques

Shape is a geometric feature which is helpful for recognition of objects. Typically, shape features are computed by numerical values without any dimensions (units) depending on as a minimum of any two shape parameters [10, 12–15]. Shape features help in understanding the morphological descriptors of objects. In this work, the shape oriented features such as area, perimeter, major and minor axes, orientation and centroid are considered. From the input microscopic images, shape-based characteristics are extracted and then passed to the classification algorithms. Then, the gender classification of *D. melanogaster* is done using different classifiers namely SVM, CNN and BPNN [38]. SVM is a supervised method, and it employs a hyperplane to discriminate the objects in different classes [38]. ANN is feed forward in nature, and it uses a BPNN. It is suitable for shape classification in a wide range of applications [39]. Here, the linear kernel SVM is used, and the model is trained based on the shape features as input data for the classification work. BPNN is utilized as a training algorithm for ANN, and input layer is fixed as about 8, and it consists of 100 masked layers, and the output layer dimension is 2.

The learning process of CNN is hierarchical in nature, and the features are generated spontaneously from the raw image datasets. Here, AlexNet architecture is used for the feature extraction process which is composed of 8 layers namely five convolution layers, two entirely linked layers and softmax as a classification layer [25]. It is essential to note that the filters are selected in the size of 11*11, 5*5, 3*3, 3*3 and 3*3 for convolution filters. In the present work, these layers are used to extract the features from the input image data and then it is passed to multi-class SVM for classification. The learning process in CNN is done according to the depth of the architecture, and lower layer generates the features like edges and corners. The mid-level layer is responsible for learning features like colors and shape, and high-level layer represents the objects present in the image. Dataset input is trained on Alexnet and the totally connected layer that is considered as a feature extraction layer. Then, the extracted deep features are passed for classification with multi-class SVM classifier. Interestingly, an average accuracy of about 95% is achieved after the successive iterative training (10 times). The generalized way of learning approach in the CNN suits well for image classification, segmentation and localization problems [25].

The performance characteristics of the presented DMG classifier are related with SVM, CNN and BPNN for the shape-based fruit fly classification tasks. From the investigation, it is inferred that the DGM attains a classification accuracy of 97 $\pm$ 1.58% followed by the CNN that achieves an average accuracy about 91.25 $\pm$ 2.38% only. BPNN employed for *D. melanogaster* gender classification achieves 88.75 $\pm$ 1.92% accuracy. Similarly, SVM classifier is also employed for *D. melanogaster* gender classification problem, and the accuracy obtained is about 87.5 $\pm$ 1.92% [38]. The DMG classifier is validated with the state-of-the-art ML methods and found that the DMG classifier outperforms with a maximum and average accuracy of about 98% and 97%, respectively. The performance of various classifiers in terms of accuracy metric is presented in Fig. 11.

The standard deviation for DMG, SVM, BPNN and CNN is also computed, and it is about 1.581, 2.061, 1.92 and 2.395, respectively. It indicates that the proposed technique is highly robust and has good generalization ability while likened with the other methods. The computational time needed to process the images by the various models are also analyzed, and it is found that DMG classifier utilizes 20 s for processing 50 images. Alternatively, CNN, SVM and BPNN have consumed 7, 5.2 and 12.7 s, respectively, for processing 50 images. Though the time is slightly high in the case of DMG, based on the nature of problem accuracy is an important metric for this research. Hence, DMG has more accuracy compared to the other methods. SVM is less time consuming in evaluation to former algorithms, but the classification accuracy is very poor. The performance of DL-based algorithms is better in terms of both time and improved accuracy when compared to NN and SVM-based algorithms. However, when the overall accuracy is considered and the time required for the classification output, the CMM model works well. Here, based on this analysis it is concluded that the proposed DMG model is good in terms of accuracy and computational time. The executable file size of the proposed DMG classifier is less than 2 KB. It is much lesser than the other comparative machine learning and deep learning models. For example, the size of executable files developed using Support Vector Machine is about 1.7 MB and AlexNet deep learning model is $\approx$ 1.2 GB. Hence, the DMG classifier is light-weight in nature, and the accuracy is high as compared to other machine learning and deep learning-based methods.

5 Conclusions

A novel DMG classifier is presented to classify the male and female of *D. melanogaster* flies from Microscopic ventral view images in an optimal time. The presented solution uses the most reliable feature (Abdomen shape) that can be obtained from a Microscopic image. Through the use of a CMM, a criterion is identified, and the boundary line is sharpened at the lower section of the abdomen. The numerical estimates in relevance to the curvature of this boundary line are also estimated with reliable methodology. The presented solution provides efficient results in the calculation of the orientation of the fly, segmentation of its body, the allocation of the boundary line of the tail and the assessment of its curvature. The classifier accuracy is found to be 98% which is 4% higher than the CNN. While compared with the accuracy of DMG with BPNN and SVM, it is found to be higher at the rate of 6% and 8%, respectively. The speed of the algorithm is about 0.4 s per image, which allows the processing of large arrays of images on a regular desktop computer.

Acknowledgements We (authors) express our sincere thanks to the National Drosophila Stock Center, Department of Studies in Zoology, University of Mysore, Mysore, for providing tools and drosophila fly data used and their support and guidance during data collection.

Funding This material is based upon work supported by Russian Foundation for Basic Research Grant under the Grant No. RFBR 20-01-00664.

Declarations

Conflict of interest There is no conflicts of interest among the author(s). This research utilizes the images of *D. melanogaster* flies, and the images are captured using the standard procedures.

References

1. Handa, J., Chandrashekara, K.T., Kashyap, K., Sageena, G., Shakarad, M.N.: Gender based disruptive selection maintains body size polymorphism in drosophila melanogaster. J. Biosci. **39**(4), 609–620 (2014). https://doi.org/10.1007/s12038-014-9452-x
2. Ahmad, F., Roy, K., O'Connor, B., Shelton, J., Dozier, G., Dworkin, I.: Fly wing biometrics using modified local binary pattern, SVMs and random forest. Int. J. Mach. Learn. Comput. **4**(3), 279–285 (2014). https://doi.org/10.7763/ijmlc.2014.v4.425
3. Neto, F.G., Rodrigues Braga, I., Harber, M.H., Paula, I.C.: *Drosophila melanogaster* gender classification based on fractal dimension. In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (2017). https://doi.org/10.1109/sibgrapi.2017.32
4. Pandey, U.B., Nichols, C.D.: Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery. Pharmacol. Rev. **63**(2), 411–436 (2011). https://doi.org/10.1124/pr.110.003293
5. Vecchio, G.: A fruit fly in the nanoworld: once again Drosophila contributes to environment and human health. Nanotoxicology **9**(2), 135–137 (2014). https://doi.org/10.3109/17435390.2014.911985
6. Bilder, D., Irvine, K.D.: Taking stock of the drosophila research ecosystem. Genetics **206**(3), 1227–1236 (2017). https://doi.org/10.1534/genetics.117.202390
7. Ugur, B., Chen, K., Bellen, H.J.: Drosophila tools and assays for the study of human diseases. Dis. Model. Mech. **9**(3), 235–244 (2016). https://doi.org/10.1242/dmm.023762
8. Dubey, T., Gorantla, N.V., Chandrashekara, K.T., Chinnathambi, S.: Photoexcited toluidine blue inhibits tau aggregation in Alzheimer's disease. ACS Omega **4**(20), 18793–18802 (2019). https://doi.org/10.1021/acsomega.9b02792
9. Roeder, T., Isermann, K., Wagner, C., Warmbold, C.: Fruit flies as models in biomedical research—a drosophila asthma model. Insect Biotechnol. (2010). https://doi.org/10.1007/978-90-481-9641-8_2
10. Priyanka, S., Sudhakar, M.S.: Geometrically modeled derivative feature descriptor aiding supervised shape retrieval. Appl. Intell. **48**(12), 4960–4975 (2018). https://doi.org/10.1007/s10489-018-1251-x
11. Abuzneid, M.A., Mahmood, A.: Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network. IEEE Access **6**, 20641–20651 (2018). https://doi.org/10.1109/access.2018.2825310
12. Tyagi, V.: Content-based image retrieval using integrated color, texture, and shape features. Content-Based Image Retr. (2017). https://doi.org/10.1007/978-981-10-6759-4_13
13. Srivastava, V., Purwar, R.K.: Classification of CT scan images of lungs using deep convolutional neural network with external shape-based features. J. Digit. Imaging **33**(1), 252–261 (2019). https://doi.org/10.1007/s10278-019-00245-9
14. Xie, Y., Zhang, J., Xia, Y., Fulham, M., Zhang, Y.: Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. Inf. Fusion **42**, 102–110 (2018). https://doi.org/10.1016/j.inffus.2017.10.005

15. Gościewska, K., Frejlichowski, D.: Classification of tooth shapes for human identification purposes—an experimental comparison of selected simple shape descriptors. Lect. Notes Comput. Sci. (2015). https://doi.org/10.1007/978-3-319-20801-5_18
16. Mallikarjuna, K., Raheem, B.A., Pathanadka, G., Suriyakumar, S.M.: A simple shape descriptor merging arithmetical wrap around technique with absolute localized pixel differences. Wirel. Pers. Commun. **117**(3), 2495–2511 (2020). https://doi.org/10.1007/s11277-020-07991-y
17. Chen, P., Li, X., Liu, J., Wu, L.: Simplifying a shape manifold as linear manifold for shape analysis. SIViP **15**(5), 1003–1010 (2020). https://doi.org/10.1007/s11760-020-01825-x
18. Ramesh, B., Xiang, C., Lee, T.H.: Shape classification using invariant features and contextual information in the bag-of-words model. Pattern Recogn. **48**(3), 894–906 (2015). https://doi.org/10.1016/j.patcog.2014.09.019
19. Huixian, J.: The analysis of plants image recognition based on deep learning and artificial neural network. IEEE Access **8**, 68828–68841 (2020). https://doi.org/10.1109/access.2020.2986946
20. Motta, D., Santos, A.Á., Winkler, I., Machado, B.A., Pereira, D.A., Cavalcanti, A.M., Fonseca, E.O., Kirchner, F., Badaró, R.: Application of convolutional neural networks for classification of adult mosquitoes in the field. PLoS ONE **14**(1), e0210829 (2019). https://doi.org/10.1371/journal.pone.0210829
21. Baker, N., Lu, H., Erlikhman, G., Kellman, P.J.: Local features and global shape information in object classification by deep convolutional neural networks. Vis. Res. **172**, 46–61 (2020). https://doi.org/10.1016/j.visres.2020.04.003
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). https://doi.org/10.1109/cvpr.2016.90
23. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50 × fewer parameters and < 0.5 MB model size (2016). arXiv:1602.07360. Top of Form
24. Barker, M., Adam: Inception V3 deep convolutional architecture for classifying acute myeloid/lymphoblastic leukemia, intel.com (2019). Retrieved 2 Feb 2019
25. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017). https://doi.org/10.1145/3065386
26. Hendry, & Chen, R.: Automatic license plate recognition via sliding-window darknet-YOLO deep learning. Image Vis. Comput. **87**, 47–56 (2019). https://doi.org/10.1016/j.imavis.2019.04.007
27. Guan, Q., Wang, Y., Ping, B., Li, D., Du, J., Qin, Y., Lu, H., Wan, X., Xiang, J.: Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J. Cancer **10**(20), 4876–4882 (2019). https://doi.org/10.7150/jca.28769
28. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation (2015). arXiv:1505.04597
29. Al-antari, M.A., Al-masni, M.A., Kim, T.: Deep learning computer-aided diagnosis for breast lesion in digital mammogram. Adv. Exp. Med. Biol. (2020). https://doi.org/10.1007/978-3-030-33128-3_4
30. Rebelo, A.R., Fagundes, J.M., Digiampietri, L.A., Bíscaro, H.H.: Methods for automatic image-based classification of winged insects using computational techniques. In: XVI Brazilian Symposium on Information Systems (2020).https://doi.org/10.1145/3411564.3411641
31. Tuda, M., Luna-Maldonado, A.I.: Image-based insect species and gender classification by trained supervised machine learning algorithms. Ecol. Inform. **60**, 101135 (2020). https://doi.org/10.1016/j.ecoinf.2020.101135
32. Li, F., Xiong, Y.: Automatic identification of butterfly species based on HoMSC and GLCMoIB. Vis. Comput. **34**(11), 1525–1533 (2017). https://doi.org/10.1007/s00371-017-1426-1
33. Chola, C., Benifa, J.V., Guru, D.S., Muaad, A.Y., Hanumanthappa, J., Al-antari, M.A., AlSalman, H., Gumaei, A.H.: Gender identification and classification of drosophila melanogaster flies using machine learning techniques. Comput. Math. Methods Med. **2022**, 1–9 (2022). https://doi.org/10.1155/2022/4593330
34. Mestetskiy, L., Bakina, I., Kurakin, A.: Hand geometry analysis by continuous skeletons. Lect. Notes Comput. Sci. (2011). https://doi.org/10.1007/978-3-642-21596-4_14
35. Sklansky, J., Chazin, R.L., Hansen, B.J.: Minimum-perimeter polygons of digitized silhouettes. IEEE Trans. Comput. **C–21**(3), 260–268 (1972). https://doi.org/10.1109/tc.1972.5008948
36. Asano, T., Toussaint, G.: Computing the geodesic center of a simple polygon. Discrete Algorithms Complex. (1987). https://doi.org/10.1016/b978-0-12-386870-1.50010-1
37. Preparata, F.P., Shamos, M.I.: Computational Geometry: An Introduction. Springer, Berlin (2012)
38. Gil-Jiménez, P., Lafuente-Arroyo, S., Maldonado-Bascón, S., Gómez-Moreno, H.: Shape classification algorithm using support vector machines for traffic sign recognition. Comput. Intell. Bioinspired Syst. (2005). https://doi.org/10.1007/11494669_107
39. Lin, S., Chou, S., Chen, S.: Irregular shapes classification by back-propagation neural networks. Int. J. Adv. Manuf. Technol. **34**(11–12), 1164–1172 (2006). https://doi.org/10.1007/s00170-006-0667-3

Leonid M. Mestetskiy earned an M.S. in Mathematics from Lomonosov Moscow State University in 1971. During 1971–1992, he worked in the Research Institute as a research fellow (1971–1975), a senior research fellow (1976–1991), and the leading research fellow (1992). In 1992–2000, he worked as a professor in Tver State University, Department of Information Systems and Technologies. In 2000, he holds a Professorship in Lomonosov Moscow State University in the Department of Mathematical Forecast Methods. He received the Ph.D. degree (1982), D.Sc. degree (1991), in Computer Science, and Professor degree (1998). His specializations are Operations research, Simulation, Computational geometry, Computer graphics and Pattern recognition. He published 120 articles and reports.

D. S. Guru is a Professor at the Department of Studies in Computer Science, University of Mysore. He completed his Doctorate in Computer Science from the University of Mysore in 2000, He was a fellow of BOYSCAT. He was a visiting research scientist at Michigan State University. He is supervising a couple of major projects sponsored by the UGC, the DST and the Government of India and has published more than 200 research papers in national and international journals and conference proceedings. He served as a Program Chair for the 2009 International Conference on Signal and Image Processing in Mysore and is currently a reviewer for the International Journal of Pattern Recognition Letters, Journal of Pattern Recognition, Journal of Document Analysis and Recognition, Journal of Image and Vision Computing and IEEE Transactions on Neural Networks.

J. V. Bibal Benifa is presently associated with Indian Institute of Information Technology, Kottayam, India as an Assistant Professor. She have obtained her B.E, M.E and Ph.D. degrees in the domain of cloud computing from Anna University, India. Her research interest are cloud computing big data analytics and machine learning. She is a life member of ISTE, India, and she is involved developing various societal applications.

H. S. Nagendraswamy obtained his M.Sc and Ph.D. degrees from University of Mysore, India in 1994 and 2007, respectively. He is currently working as Associate Professor in the Department of Studies in Computer Science, University of Mysore, Manasagangothri, Mysore, Karnataka, India. His focused areas of research include Shape analysis, Texture analysis, Sign Language Recognition, Precision agriculture, Symbolic data analysis, Fuzzy theory, Biometrics and Video analysis. He has been a reviewer for many international journals including Fuzzy system, Pattern Recognition.

Channabasava Chola is presently associated with Indian Institute of Information Technology, Kottayam, India as Full Time Ph.D. Research Scholar. He have obtained his BCA and MCA degrees in the domain of Computer Applications from Mysore University, India. His research interests are image processing and machine learning.