Journal of Artificial Intelligence in Medical Applications (JAIMA)
Vol. 1, No. 1, 2025, **pp**. 135–146
ISSN: 3106-1494 | **DOI**: https://doi.org/10.65511/jaima.v1i1.744

Automatic Detection of Landmark Points for Estimating the Brain's Dimensions of the Mouse in X-ray Images Using Machine Learning Approache

Omar Al Okashi[1,*], Hongbo Du[2], Hisham Al-Assam[3]

[1]Department of Computer Networks Systems, College of Computer Science and Information Technology , University Of Anbar, Ramadi, Iraq, omar.alokashi@uoanbar.edu.iq
[2]Department of Applied Computing, College of Computing, The University of Buckingham, Buckingham, United Kingdom, hongbo.du @buckingham.ac.uk
[3]Department of Applied Computing, College of Computing, The University of Buckingham, Buckingham, United Kingdom, hisham.al-assam@buckingham.ac.uk

Received: 25/09/2025, Revised: 19/11/2025, Accepted: 28/11/2025, Published: 30/12/2025

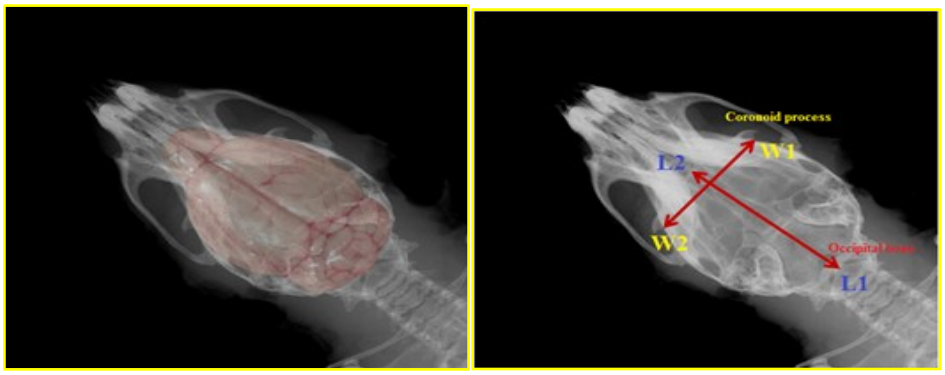
ABSTRACT: Measure of brain dimensions is an important metric for discovering brain abnormalities associated with certain diseases. Using an animal model to measure brain dimensions, followed by a discovery of possible genes responsible for changes in the dimensions, enables better understanding of the underlying genetic cause of the brain's abnormalities in humans. This paper presents a three-stage automatic solution for estimating brain dimensions from dorsal-ventral X-ray images of mouse heads based on machine learning techniques. At the first stage, a skull modelling technique based on circular Hough transform is used to determine region of interests of the key landmarks on the skull. In stage 2, the procedures for the width and length landmarks ROIs estimation based on a circle model are presented. The landmark locations are then obtained in stage 3 using the random forests classification method based on texture features derived from the local binary patterns and histogram of oriented gradients and then brain dimensions are measured in terms of relevant width and length of the skull. Experimental results on a set of X-ray head images of 798 wild type and 6,090 mutant mice have demonstrated the effectiveness of the proposed solution in separating images of abnormal brain dimensions from those of normal ones. Furthermore, the estimated dimensions have also led to identifying 7 possible genes responsible for changes in the brain size of the mutant mice, some of which were already linked to Down's syndromes, which further demonstrates the practical usefulness of the proposed solution.

Keywords: Brain, X-ray, Landmark, Skull, Gene.

1. INTRODUCTION

The recent development of high-throughput production of targeted, genetically-altered mouse lines has enabled a systematic phenotypic screening and gene discovery [1]. Identifying genes correlated with changes in brain size is a key step towards understanding the genetic basis of neurological and developmental brain syndromes. Accurate measurement of any changes in brain size is of a significant biological and clinical relevance. For example, mutations in the Mcph1 gene in human have been associated with Microcephaly, a disease defined as a head circumference is more than three standard deviations below the age-related mean [2], which is often associated with incomplete brain development [3]. In contrast, Macrocephaly has been reported in patients with mutations in the Kptn gene, which is also characterised by neurological disorders [4]. It is likely that there are further unidentified genes in which mutations may also cause neurological or behavioural symptoms relating to alteration of brain dimensions. Therefore, the identification of genes associated with dimensions' alterations may provide insight into both specific and generalised abnormalities with the brain.

Human annotation is often used to measure brain dimensions from the skull outline of mouse X-ray images as illustrated in Figure1. Figure 1(a) shows the actual location of the brain in the mouse's head. The trained analyst firstly locates four key landmark points, i.e. W1, W2, L1 and L2, on the mouse skull image as shown in Figure 1(b). The dimensions of the brain, i.e. brain width and length, are then measured as the distance between W1 and W2 and the distance between L1 and L2 respectively. It has been found that the targeted position for W1 or W2 is the cross point between the skull edge line and the root of the coronoid process of the mandible of the mouse [5]. For the length landmarks, L1 is located at the base of the occipital bone while L2 is located on Nasalis – midline junction of nasal and frontal suture- as shown in Figure 1(b).



(a) (b)

Figure 1. Manual Annotation of mouse brain dimension landmark points, adapted from [5] (a) The location of actual brain image of mouse in relative to the skull and (b) Width and length landmark points manually annotated on the mouse's head X-ray image

Automatic size estimation of a 3D object such as the brain from a 2D X-ray image is a technically challenging task due to difficulties in locating related and reliable landmarks within the image caused by poor contrast between intensity values for different tissues such as bone, brain, vessels, etc., and overlapping of pixel values representing the third dimension of the organs. In addition, specific challenges also exist for locating width and length landmark points correctly. The variations of size and layout of the coronoid processes across different images add a level of difficulty for locating W1 and W2. Although L1 landmark is located on the base of the occipital bone, the base edge of this bone is hard to locate due to extreme poor contrast in the area. The L2 landmark is even more difficult to locate for the same reason. Furthermore, the highly varying and inconsistent structures of the bones and tissues in the neighbouring areas for both L1 and L2 landmarks across different X-ray images of mouse's heads make the location of both landmarks very difficult. Individual differences of size of the mouse's head from image to image add another level of difficulty for fully automated solutions. All the aforementioned challenges are even felt by trained expert analysts, causing inter- and intra-analyst variations in their manual measurement results. This apparent lack of a solid ground truth will lead to difficulties in objectively evaluating automated solutions. Given the volume of images that are generated by high-throughput screening processes, an automated solution for annotation and triaging of images for human analysis will not only reduce costs in time and human resources, but also eliminates the likelihood of inconsistency in manual measurements.

In the literature, there are several publications to label and segment the brain of the mouse manually, semi-automatic or in a full automatic manner. Most of these proposals were dedicated to 3D MRI images because the ability of such imaging modality to recognize the soft tissues of the brain. Some manual and semi-automatic methods were used for labelling mouse brains ([6], [7] and [8]). However, the manual and semi-automatic methods are considered as time-consuming and error prone. Segmentation of the mouse brain from 3D MRI images into different stuctures was propsed in several proposals ([9], [10] and [11]). For those methods, the registiration was firstly done using an atlas based on image intensities matching. After that, the segmentation of the brain from the image was achieved based on wraped label information about the structures. The important note about those methods is the use of single atlas to perform image registiration which has a drawbacks of bias existence that might affect the segmentation [12].

2. METHOD

The proposed method attempts to automate the estimation of the brain dimensions by locating the same four key landmark points as trained analysts do (see Section 1). In principle, the proposed method is based on accurate modelling of the skull. This skull model will then be used for detecting region of interests (ROIs) of the landmarks. Another important feature of the proposed method is to adopt a supervised learning approach in locating the optimal final positions of the landmarks. As shown in Figure 2, the method consists of the following three key steps:

1. Modelling the skull by the best fit circle. The circle model was chosen due to its suitability to approximate the initial positions of all four key landmarks.
2. Detecting the landmarks ROIs for the key landmarks based on the circle model.
3. Locating the landmark positions and estimating the brain dimensions. Landmarks detection is done based on supervised machine learning based approach.

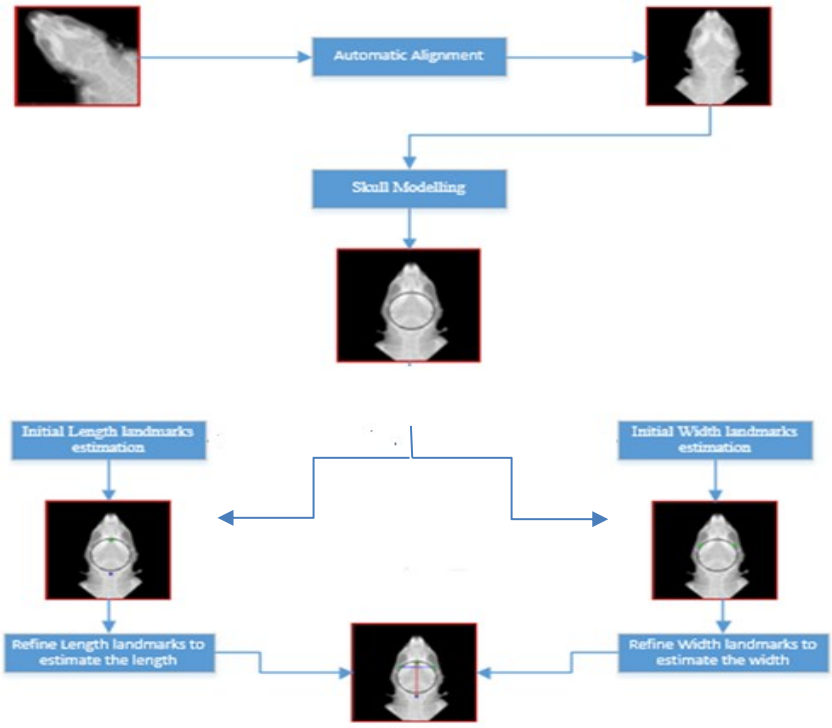
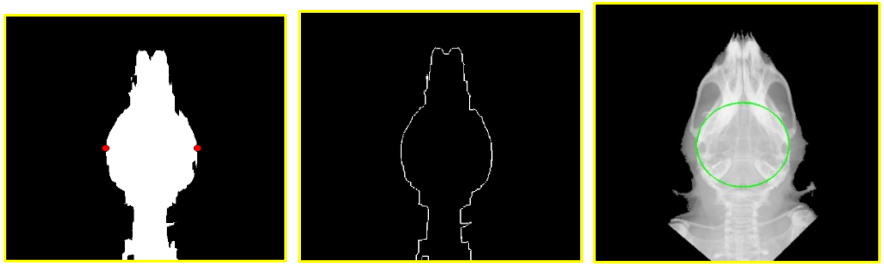


Figure 2. The main stages of the proposed method

2.1. Skull modelling

It could be seen from the Figure 1(b) that three of the four landmarks are in fact located on the edge of the skull while the other landmark is above the hypothetical width line formed by two of the three landmarks. This structural property of the mouse skull motivated us to model the skull with a simple regular geometrical shape that helps to find the skull edge and hence the approximate initial locations for the landmarks. Circle appears a good choice for its closeness to the skull shape and its ease to derive the initial locations.

The first step to locate the best fit circle is to prepare a suitable skull mask as shown in Figure 3(a). The initial radius of the circle is then estimated by the half of the distance between the leftmost and the rightmost points (red points) of the skull mask, as shown in Figure 4(a). To prepare a skull model, Hough Transform (HT) is applied to find the best fit circle of the skull mask. The HT [13] is a well-known transform that has an ability to locate geometrical shapes in an image [14]. The reason behind this choice is the powerful ability of HT to detect the circle with known parameters from a binary image. The input parameters to HT function are the outer border of the skull mask (shown in Figure 3(b)) which obtained by using Canny edge operator [15] and the candidate radiuses. Due to the importance of the best fit circle, as a step of caution, the initial radius and six other radiuses (±3 pixels of initial radius) are treated as candidate radiuses. The detected circle of the skull is as shown in Figure 3(c) after superimposing it on the aligned input image.



(a) (b) (c)

Figure 3. Skull modelling (a) Rotated skull mask and the right and left most points; (b) Canny edge of the skull mask and (c) Skull modelling using a circle

2.2. Initial detection of landmarks ROIs

In this section, the procedures for the width and length landmarks ROIs estimation based on a circle model are presented. Determination of ROIs of landmarks are very important to make the final refined positions more accurate and efficient.

Width Landmarks ROI Determination

From the skull modelling, two observations that influence the estimation of the width landmarks ROIs are made. The first observation is that the detected circle, when it is superimposed on the aligned image, should pass through the edge of the skull near to the coronoid processes. The second observation is that the coronoid processes in the aligned image have almost the same or very close y-coordinate. Based on these two observations, two initial seed points are found on the circumference of the circle using the following formula to obtain x and y coordinates for every point:

$$x = r \times cos(t) + x_c \quad (2)$$

$$y = r \times sin(t) + y_c \quad (3)$$

where r is the radius of the detected circle, x_c and y_c are the x and y coordinates of the centre of the circle, and t is an estimated angle. To locate the initial seed points, the angle t is set to 45º for the right side and 135° for the left side respectively. These specific angles are selected from the observation of the position of the coronoid processes relative to the centre of the circle. The initial width landmarks detected points and correspondence width landmarks ROIs that include coronoid processes are shown by red boxes in Figure 7.

Lower Length Landmarks ROI Determination

The ideal L1 landmark point should be located on the lower edge of the occipital bone as shown in Figure 1(b). The lowest point of the circle model was developed in Section 2.2, sits either above, below or on the bone itself which makes it unsuitable to be directly considered as L1 initial location. Find a consistent initial location for L1 determines the final accurate ROI for L1 landmark, and hence must be chosen with care. To locate the ideal initial location for L1, we use similar procedure as used in Section 2.2 to determine the upper edge of the occipital bone. From this point, the ROI that contain the target landmark could be estimated as shown by blue box in Figure 4.

Upper Length Landmarks ROI Determination

By the observation, the required upper landmark point is always located near the uppermost point of the detected circle. For this reason, the topmost point of the circle is used to estimate the ROI for upper length landmark as shown by blue box in Figure 4.

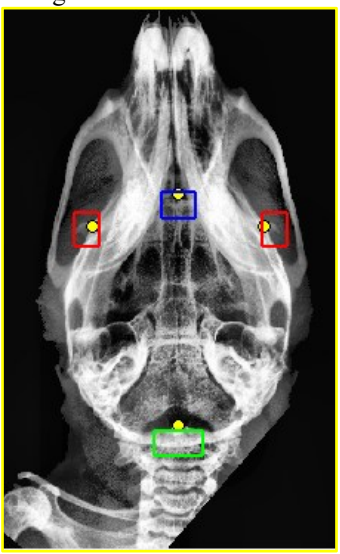


Figure 4. Landmarks ROIs

2.4. Landmarks detection based on random forests

This section proposes the use of a machine learning approach based on Random Forests (RF) to detect the final location for the width and length landmarks. RFs are an ensemble learning method based on generating multiple tree predictors where each tree in the forest depends on a random vector of features sampled independently [16]. Because of the limitation of intensity based approaches to detect different landmarks from medical X-ray images, the decision was made to use the RF approach along with different texture features for landmark detection. In addition, the using of RF approach for landmark detection from X-ray images showed satisfactory results recently. The approach mainly goes through two stages, which are

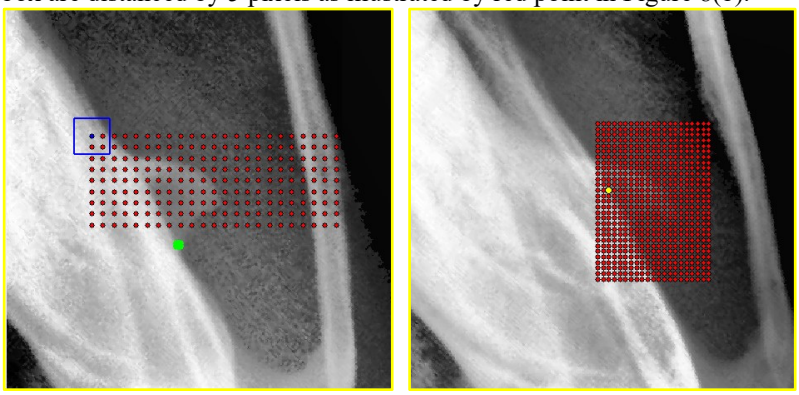
training the model of classification exploiting manually annotated images and then predicting the landmark points in new unseen images.

At the training stage, the proposal is to train two separate models - RF with 500 classification trees - for each landmark point using enhanced image as in Figure 4. The enhanced image using Contrast Limited Adaptive Histogram Equalization (CLAHE) [17] shows the better contrast of the different landmarks ROIs than the input image.

For the width landmarks detection, the annotator selects only one pixel that corresponds to each landmark point in the manual annotation process. In practice, the precision of the manual landmark selection is not at a pixel level i.e. any of the surrounding pixels within a window of size (D × D) can be considered as a correct landmark as shown by a green window in a zoom-in image in Figure 7(a). The number of positive samples is then increased to (Number of training images × D2). For the negative examples (non-landmark points), from every training image, a group of pixels are sampled automatically from windows with dimensions of (x × y) pixels which located 15 pixels above of the manually annotated landmark point. Because of similar features are expected from adjacent pixels, the selected negative samples are distanced by 10 pixels as shown by red points in Figure 7(a). The position and the dimensions of the negative samples window are selected to cover all possible types of different tissues and structure expected in ROI for landmark in the prediction stage.

Every positive and negative sample pixel are then described by two types of feature values, LBP and HOG, which extracted from a square patch of size (33 × 33) pixels which centred at the sample pixel itself. Figure 8(a) shows the patch around the first negative sample in blue. Those two feature descriptors produce their features in a different manner. The LBP descriptor is powerful in recognizing monotonic grey scale changes like illumination variations [18]. Such changes exist in the ROI landmark especially in the transition area between the coronoid process and the background or the skull border. On the other hand, the HOG descriptor is focused on the orientations change in a set of pixels [19]. In the width landmark area, the orientation changes associated with the edges of the coronoid process and the skull.

The features collected from every patch associated with label of sample (0 or 1) are used to train the RF model (RF-LBP and RF-HOG) to predict more likely position for the landmark point in a prediction stage. To improve the efficiency of the proposed solution, a box the of size (j × k) pixels around the initial location of W1 and W2 landmarks as shown in Figure 6 is used at the test stage instead of scanning the whole image. To furthermore decrease the time consuming and preserve the accuracy level, the selected test samples from the box are distanced by 5 pixels as illustrated by red point in Figure 8(b).



(a) (b)

Figure. 5. Selected pixels in training and prediction stages (a) Positive (green points) and negative samples (red points) sampled in the training stage and the patch of the first sample(b) Pixels sampled from W1 ROI in a prediction stage

For each selected pixel in a prediction window, two types of features are extracted of its patch exactly as mentioned in the training stage. After that, each of pixels in the prediction area is predicted by both RFs models obtained in the training stage to determine its class. Each tree in each model will candidate number of pixels as true positive pixels. The candidate landmark point for a specific model (RF-LBP and RF-HOG) is selected as the pixel that collect the highest number of votes from the trees in the model. Similarly, the final location of the landmark will be the pixel that receive the highest number of votes from the trees in both RF-LBP and RF-HOG models (RF-LBP-HOG).

The RF classification procedure is proposed to detect the L1 and L2 landmark points from their ROIs. The general approach application is similar to what has been applied in width landmark estimation but with some changes in parameters setting to suit the different in landmarks ROI. In the training stage, the positive samples (landmark points) are again extended by considering the annotated point by the expert annotator and any of surrounding pixels within a window of size (D × K) for L1 and L2 landmarks. For both

landmark points, the number of added positive samples in the horizontal axis (K) is more than the case with width points. This because the focus here is mainly on the vertical position rather than the horizontal one, so some tilting to the left or right will not affect the accuracy of the final length estimation. Also, because the search areas of annotated L1 and L2 landmark points have a lot of different structures, the negative samples (non-landmarks point) are sampled from two windows to above and below annotated L1 and L2. The training for two models using the two extracted types of features (LBP and HOG) is done exactly the same as the training for the width landmark points. In the testing stage, for each unseen image, pixels are sampled from the ROIs and then sent to the model to predict whether it positive or negative examples.

3. RESULTS AND DISCUSSION

This section evaluates and reports the results of applying the proposed method on the dataset. Firstly, a description of the materials used to evaluate the method are given. After that we introduce the experimental results for manual versus automatic evaluation in addition to the identification of candidate genes. Finally, a discussion of the method and results is presented.

1- Materials

The data set used for evaluating the proposed method contains the close-up head dorsoventral X-ray greyscale images of 256 levels of intensity and of the size of 2368 × 2340 pixels. The data set is divided into two subsets. The first subset includes 798 images of wild type (WT) mice of normal mouse skeleton. The second subset contains 6,090 images of mutant mice obtained from 435 mutant lines where each mutant line has 14 images (7 males and 7 females). The data sets are provided by the Wellcome Trust Sanger Institute (WTSI), Sanger MGP [5]. The data sets are publicly available for research use under terms and conditions that we follow for this study.

To produce manual measurements for evaluation purposes, the manual process as explained in Section 1 was followed by an expert from WTSI on a random sample of 100 X-ray images.

2- Experimental Results

In this part, the evaluation of manual versus automatic measures and the gene identification are presented.

a- Manual versus automatic measurements

Due to the large volume of images and the infeasibility of manually locating all landmark points on all images, random samples of 100 images were annotated by an expert to be used as a ground truth for evaluation purposes. To assess the accuracy of the detection of the four landmark points, the Euclidian distances between the automatically detected landmark points and the corresponding ground truth ones were calculated in millimetres (mm). The scale for the conversion from pixel to millimetre for the mouse skull image adopted in this research is 0.012 mm for every pixel [5].The mean (μ) and the standard deviation (σ) of Euclidian distances between the obtained and ground truth landmarks were calculated and reported in Table 1.

Table 1. Results of all landmark points obtained using 50 trees models in terms of mean and standard deviation of the Euclidian distances between the automated landmark points and the ground truth and the SDR within 0.1mm, 0.25 and 0.5mm neighbourhoods

	RF-LBP				RF-HOG				RF-LBP-HOG			
	μ	0.1 mm	0.25 mm	0.5 Mm	μ	0.1 mm	0.25 mm	0.5 mm	μ	0.1 mm	0.25 mm	0.5 mm
W1	0.20 ± 0.16	39	69	95	0.26 ± 0.27	35	61	84	0.21 ± 0.22	41	70	91
W2	0.15 ± 0.11	41	87	97	0.14 ± 0.15	52	85	97	0.13 ± 0.12	54	89	98
L1	0.28 ± 0.20	14	56	86	0.27 ± 0.19	15	60	85	0.28 ± 0.18	13	56	87
L2	0.35 ± 0.20	5	38	81	0.33 ± 0.22	16	44	80	0.31 ± 0.20	8	46	84
Total	0.24 ± 0.17	25	63	90	0.25 ± 0.21	30	63	87	0.23 ± 0.18	29	65	90

The second evaluation method was used is the success detection rate (SDR) with respect to the 0.1mm, 0.25mm and 0.5mm precision range. The SDR was expressed in terms of percentage of automatically obtained landmarks located within predefined range of ground truth landmarks as shown in Table 1.

This table gave us a good indication about the accuracy obtained by different proposed RF models. For the width's landmark points, the first note could be made from the above table that all models detected W2 (left landmark) with higher accuracy than W1 (right landmark). To understand this situation, a close examination of the results showed that the number of difficult cases (i.e. small or not recognized coronoid processes) in the right side are more than the left ones. This note could also explain why the HOG-model respond better than the LBP-model for the W2 (especially for 0.1 mm SDR) when the coronoid processes and thus the orientation information are more apparent. However, the combining between two models could add noticeable enhancements for both landmarks and especially for W2. Another note could be made from Table 1 is that the result of width landmarks detection is better than the length one which reflect the difficulty of length landmark detection.

On the other hand, the nature of difficulties in locating the length landmarks is different from the width one. It is clear that both RF-HOG, RF-LBP and RF-LBP-HOG models have approximately similar accuracy for L1 landmark detection. The high variations of the texture around the L2 landmark from image to image reduces the ability of RF model to learn the L2 landmark which leads to poor results in comparing with other landmarks detections. However, RF-HOG model still outperforms the RF-LBP model for L2 landmark detection. For the total, RF-HOG model shows slightly better result than RF-LBP in terms of SDR for 0.1mm. on the other hand, RF-LBP shows slightly better mean and standard deviation of Euclidian distances than RF-HOG. The good thing that the RF-LBP-HOG can give the best of the two models in terms of SDR and mean and standard deviation.

The results reported on the previous table were based on using 50 classification trees per each model. However, to draw a conclusion on the effect of increasing the number of trees on the accuracy, we repeated the experiment with models of 500 trees and the results are reported in Table 2.

Table 2 Results of all landmark points obtained using 500 trees models in terms of mean and standard deviation of the Euclidian distances between the automated landmark points and the ground truth and the SDR within 0.1mm, 0.25 and 0.5mm neighbourhoods

	RF-LBP				RF-HOG				RF-LBP-HOG			
	μ	0.1 mm	0.25 mm	0.5 mm	μ	0.1 mm	0.25 mm	0.5 mm	μ	0.1 mm	0.25 mm	0.5 mm
W1	0.23 ± 0.20	32	64	91	0.25 ± 0.26	41	63	86	0.22 ± 0.23	41	65	90
W2	0.15 ± 0.16	48	85	95	0.15 ± 0.16	55	85	95	0.12 ± 0.11	61	92	96
L1	0.27 ± 0.18	17	55	88	0.26 ± 0.17	18	59	90	0.26 ± 0.18	16	59	88
L2	0.32 ± 0.20	8	48	84	0.31 ± 0.22	15	49	82	0.29 ± 0.19	13	51	87
Total	0.24 ± 0.18	26	63	90	0.24 ± 0.20	32	64	88	0.22 ± 0.18	33	67	90

For all the landmarks, the results confirmed the conclusions made from Table 1. However, we can see that the results in terms of SDR and mean and standard deviation of Euclidean distances has been enhanced for most of the cases.

After assessing the three models to detect the different landmarks, one needs to evaluate the proposed width and length of the brain based on each of three models versus the manual measurements taken by a domain expert. To perform such evaluation, we decided to use Bland-Altman analysis to estimate the agreement between the results of the manual and automatic methods [20]. This selection is motivated by the fact that using other popular methods such as correlation coefficients might be misleading. The idea behind the Bland-Altman analysis is mainly based on plot the difference between the measurements of two methods against their mean. This should be followed by calculating the bias (mean of differences $\bar{d}$) and the standard deviation (σ) of the differences to estimate the lower and upper limit of agreement (LOA). For the normally distributed differences, 95% of differences lie between the lower LOA ($\bar{d}$ - 1.96σ) and Upper LOA ($\bar{d}$ + 1.96σ).

Figure 6 shows the scatter plot for the Bland-Altman analysis for brain's width and length respectively when using RF-LBP-HOG model with 50 and 500 trees to estimate the measurements.

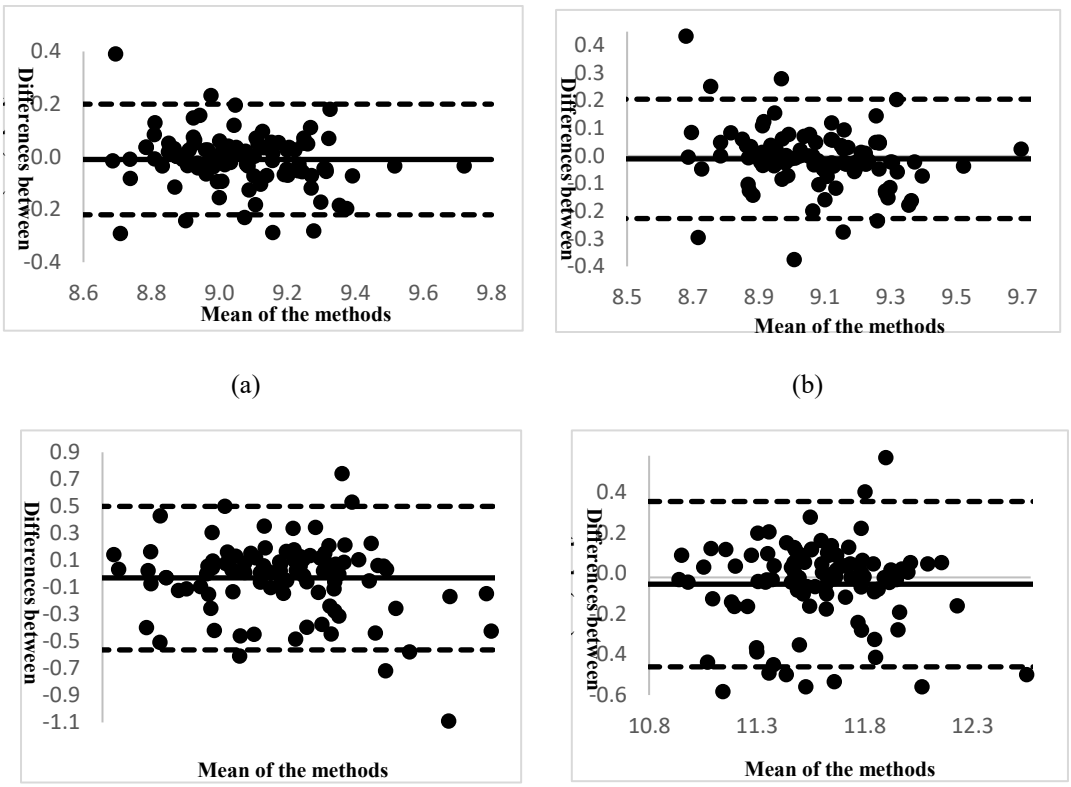


Figure 6. Scatter plot of Bland-Altman analysis for the automatic width and length measurements obtained from RF-LBP-HOG model versus manual measurement (a) Automatic width obtained with 50 trees model (b) Automatic width obtained with 500 trees model (c) Automatic Length obtained with 50 trees model and (d) Automatic Length obtained with 500 trees model

Tables 3 and 4 show the upper and lower LOA and the bias for the RF-LBP, RF-HOG and the RF-LBP-HOG models for width and length respectively.

Table 3. LOAs and Bias obtained from applying Bland-Altman analysis on the width measurement

Measures (Width)		Upper LOA	Lower LOA	Bias (Mean of differences)
RF_LBP	50 Trees	0.22	-0.24	-0.01
	500 Trees	0.31	-0.31	0
RF_HOG	50 Trees	0.30	-0.30	0
	500 Trees	0.25	-0.26	-0.01
RF_LBP_HOG	50 Trees	0.21	-0.21	0
	500 Trees	0.21	-0.23	-0.01

Table 4. LOAs and Bias obtained from applying Bland-Altman analysis on the Length measurement

Measures (Length)		Upper LOA	Lower LOA	Bias (Mean of differences)
RF_LBP	50 Trees	0.59	-0.66	-0.03
	500 Trees	0.52	-0.50	-0.01
RF_HOG	50 Trees	0.47	-0.68	-0.11
	500 Trees	0.46	-0.59	-0.06
RF_LBP_HOG	50 Trees	0.48	-0.58	-0.05
	500 Trees	0.37	-0.44	-0.03

Several conclusions could be made from the figures and the tables above. Firstly, the closeness between the manual and the automatic estimation is good in most cases. However, the width still shows more consistency than the length due to the underlying difficulty in in locating the length landmarks as explained

earlier. For the width, the LOA is better when using models with 50 trees than using 500 trees and at the same time the LOA for the RF-LBP-HOG model outperforms the other two models.

For the length estimation, there are a noticeable enhancement for the LOA when using the 500 trees model. This is expected because of the improvement achieved with the length landmark detection with 500 trees model as explained in Table 1. Similar for the width, the LOA for the RF-LBP-HOG model outperforms the other two models. We can conclude that the more difficult landmarks such as length one may receive more improvement when increasing the number of trees than the width landmarks.

b- Genes Identification

The ultimate goal of the experiment is to identify the genetic basis of abnormalities in the brain size i.e. what are the genes that could lead to a bigger or smaller brains?. To do this, one first needs to establish what is normal and what is not when it comes to brain size of mice for both males and females separately. This is typically done based on the RR approach to identify the LB and UB of the width and length measurements obtained from WT animals. To robustly and reliably identify the genes that are associated with changes in a brain size for the mice mutant group, the Reference Range (RR) methodology based on percentiles is followed [1] to establish the normal range of WT variations. The RR is applied separately on the male and female mice to consider the gender effect. The RR is established from the measurements obtained from applying the automated method on the 798 WT mice (normal) images. The Lower Bound (LB) and the Upper Bound (UB) of the 95% confidence interval of the WT measurements are set at the 2.5 and 97.5 percentile values respectively [1]. After that, the process of identifying the genes associated with abnormalities in brain size goes through the following two stages.

Stage1:

Each width and length measurement of every mutant mouse was compared against the relevant LB and UB to establish whether the measurement falls within or outside the RR. For example, Figure 10 shows the distribution of the brain width estimation of the male for 3045 mutant mice. The vertical green lines represent the LB and UB of the RR which obtained from the width measures of WT male (normal) cases.

Out of 6090 mutant mice images analysed, 5133 (84.3%) were found to be within the RR established by the WT measurements, and 957 (515 male and 442 female mice) (15.7%) of mutants fell outside the RR for both the width and the length measurements. However, it is important to highlight that not all samples that fall outside the RR are necessarily true hits (abnormal cases). Arguably the above number includes many false positive cases due to two main reasons. First is related to any incorrect landmarks detection which leads to incorrect dimension estimation. Second reason is that although the case is correctly measured as abnormal size but this case might belong to a specific gene where the total number of cases of abnormal size for it is less than the hit rate threshold (to be explained next in Stage2).

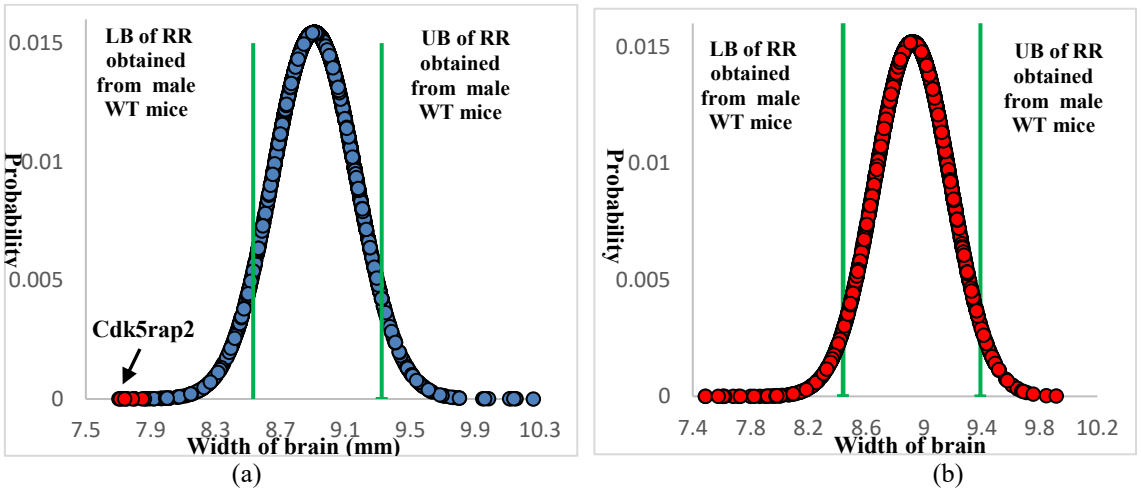


Figure 7. Normal distribution of width estimation for mutant population with RR bound
(a) Male mice and (b) Female mice

Stage2:

To filter out the false positive cases, we calculated the so-called hit ratio of each mutant line, which represents the percentage of images that falls outside the RR out of the total number of images in the mutant line (i.e. how many animals out of 7 per mutant line per gender in our case?). For example, the hit rate of a mutant line is 100% if all the 7 images of that line fall outside the RR. In this research, we set the hit rate threshold at 60% i.e. to consider a mutant line as a true hit, at least 5 out of 7 mice must fall outside the RR

[1]. Out of the 957 outlier cases, only 77 cases were considered as true hits as shown in Table 1 where the "-" mark is used to indicate that the ratio is below 60%.

Table 5. Possible gene hits

Name of allele	Gender	Width of Brain		Length of Brain	
		Phenotype	Hit Ratio	Phenotype	Hit Ratio
Arl4d/Arl4d	Male	<LB	83%	Normal	-
	Female	<LB	71%	<LB	100%
Cdk5rap2/Cdk5rap2	Male	<LB	100%	Normal	-
	Female	<LB	100%	<LB	80%
Mcph1/Mcph1	Male	<LB	66%	Normal	-
	Female	Normal	-	<LB	66%
Agap1/Agap1	Male	Normal	-	<LB	71%
	Female	Normal	-	<LB	66%
Pthlh/+	Female	Normal	-	<LB	66%
Aspm/Aspm	Male	<LB	66%	Normal	-
NovelGeneX	Male	Normal	-	<LB	66%

The table above shows a list of 7 genes that could be potentially responsible for abnormalities in the brain size. Most of listed genes are already confirmed or investigated in the literature to have a certain relationship to the change in the skull shape and thus the brain size [21]. This in a way confirms the effectiveness of the developed method in identify the list of genes responsible for changes in the brain size. Interestingly, the table shows a possible novel gene being temporarily called 'NovelGeneX', which is currently under investigation by the WTSI team. For illustration purposes, the instances of one gene (Cdk5rap2) that confirmed to have smaller width of brain are highlighted as red dots in Figure 10(a).

3- Discussion

The list of 7 genes along with the manual vs. automatic comparison presented in the previous section illustrates that the proposed method is effective in estimating brain's dimensions. Section 2.3 highlighted the fact that the correct skull modelling is a key step towards accurate dimensions' estimation. To increase the overall accuracy of applying the proposed method, a reliability criteria proposed in Section 4.3.2 was used to exclude a small percentage (0.28%) of images which cannot be reliably segmented. Out of 6891 images were tested, the proposed method succeeded in identifying inaccurate circle positions for 185 images of which 166 cases were correctly refined by the correction procedure in the skull modelling explained in Section 2.4. Only 19 images were finally rejected as images of "unsatisfactory quality". In other words, the proposal produced automatic measurement on 99.72% of the total number of images. We showed that RF classification approach results depends on the type of features used to train the model. To exploit this, we produced three different estimation for each landmark based on the HOG, LBP and the merging between the HOG and LBP features.

To sum up, even with some small inaccuracy in a handful of measurements, the proposed method is considered as helpful tool for manual annotator. Interestingly, it was able to detect changes in the brain dimensions associated with mutation in the genes Mcph1 and Cdk5rap2 [21], which are known genes for microcephaly. Giving the fact that only the abnormal measurements are worthy for manual assessment, the automated method minimizes the number of images to be checked manually dramatically, which leads to focussing the attention on more likely candidate genes.

4. CONCLUSION

Accurate estimation of the brain size from high-throughput X-ray images of mutant mice is essential for understanding the genetic basis of brain abnormalities. As shown at the introduction section, our review of relevant literature informed us that our research work is (at least as far as we know) the first original work on mouse X-ray image, that the problem of landmark detection in medical and biomedical images is still an open problem where a one-fits-all solution does not exist, and that supervised machine learning solutions can assist landmark detection.

This paper described a fully-automated algorithm that models the skull and locates different landmark points to estimate the width and length of the brain in X-ray images. The proposal exploited machine learning based techniques to locate the accurate final locations of the landmarks which directly influence the accuracy of brain dimension estimates. The estimation results produced by the proposed method succeeded in identifying a list of genes that might be associated with the changes in the brain

dimensions. Some of these genes were already known to cause Microcephaly, which demonstrates the effectiveness of the proposed solution. At the same time, the proposed method helped in isolating a potentially interesting unknown gene (NovelGeneX), further indicating the potential of the proposed method. Furthermore, the evaluation in terms of the manual versus automatic measurements illustrated it viability. The paper finally argues that the automation of landmarks from biomedical images has huge potentials in reducing human errors and cutting costs.

ACKNOWLEDGEMENTS: The authors express their gratitude to the Wellcome Trust Sanger Institute Cambridge and Mr Chris Lelliott and Mr Simon Maguire in particular for providing the test images for the research and useful discussions.

FUNDING INFORMATION: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORS CONTRIBUATIONS: [Author 1]: Conceptualization, methodology, , writing – original draft.
[Author 2]: Data analysis, visualization, writing – review & editing.
[Author 3]: Supervision, data collection, project administration, validation. All authors read and approved the final manuscript.

COFLICTS OF INTERESTS: The authors declare that they have no conflicts of interest.

EITHICAL APPROVAL: Ethical approval for this study was obtained from [Wellcome Trust Sanger Institute Cambridge].

DATA AVAILABILITY STATEMENTS: The data supporting the findings of this study are available from [https://www.sanger.ac.uk/data/mouse-genomes-project].

REFERENCES

[1] J. K. White, A.-K. Gerdin, . K. A. Natasha , E. Ryder, M. Buljan, J. N. Bussell, J. Salisbury, S. Clare, N. J. Ingham, C. Podrini, R. Houghton, J. Estabel, J. R. Bottomley, D. G. Melvin, D. Sunter, N. C. Adams, D. Tannahill, D. W. Logan, D. G. MacArthur, J. Flint, V. B. Mahajan, S. H. Tsang, I. Smyth, F. M. Watt, W. C. Skarnes, G. Dougan, D. J. Adams, R. Ramirez-Solis, A. Bradley e K. P. Steel, "Genome-wide Generation and Systematic Phenotyping of Knockout Mice Reveals New Roles for Many Genes," *Cell,* vol. 154, pp. 452-464, 2013.

[2] A. Jackson, H. Eastwood, S. . M. Bell, . J. Adu, C. Toomes, I. Carr, E. Roberts, D. Hampshire, Y. Crow, A. Mighell, G. Karbani, H. Jafri, Y. Rashid, R. Mueller, A. Markham e C. Woods, "Identification of microcephalin, a protein implicated in determining the size of the human brain," *The American Journal of Human Genetics,* vol. 71, nº 1, pp. 136-142, 2002.

[3] J. Ross e J. Frias, "Microcephaly. In: Congenital malformations of the brain and skull Part 1," *Elsevier Holland Biomedical Press,* vol. 30, pp. 507-524, 1977.

[4] E. Baple, R. Maroofian, B. Chioza, M. Izadi, H. E. Cross, S. Al-Turki, K. Barwick, A. Skrzypiec, R. Pawlak, K. Wagner, R. Coblentz, T. Zainy, M. A. Patton, S. Mansour, P. Rich, B. Qualmann, M. E. Hurles, M. M. Kessels e A. H. Crosby, "Mutations in KPTN cause macrocephaly, neurodevelopmental delay, and seizures," *The American Journal of Human Genetics,* vol. 94, nº 1, pp. 87-94, 2014.

[5] W. T. Sanger Institute, "Bone and X-ray Analysis Basics," 2013.

[6] A. Badea, P. Nicholls, G. Johnson e W. Wetsel, "Neuroanatomical phenotypes in the reeler mouse," *NeuroImage,* vol. 34, nº 4, pp. 1363-1374, 2007.

[7] J. Redwine, B. Kosofsky, R. Jacobs, D. Games, J. Reilly, J. Morrison, W. Young e F. Bloom, "Dentate gyrus volume is reduced before onset of plaque formation in PDAPP mice: a magnetic resonance microscopy and stereologic analysis," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, nº 3, pp. 1381-1386, 2003.

[8] J. Zhang, Q. Peng, Q. Li, N. Jahanshad, Z. Hou, M. Jiang, N. Masuda, D. R. Langbehn, M. I. Miller, S. Mori, C. A. Ross e W. Duan, "Longitudinal characterization of brain atrophy of a Huntington's disease mouse model by automated morphological analyses of magnetic resonance images," *NeuroImage,* vol. 49, nº 3, pp. 2340-2351, 2010.

[9] A. Ali, A. Dale, A. Badea e G. Johnson, "Automated segmentation of neuroanatomical structures in multispectral MR microscopy of the mouse brain," *NeuroImage,* vol. 27, nº 2, pp. 425-435, 2005.

[10] J. Lee, J. Jomier, S. Aylward, M. Tyszka, S. Moy, J. Lauder e M. Styner, "Evaluation of Atlas based Mouse Brain Segmentation," em *Proceesings of SPIE*, 2009.

[11] A. Sharief, A. Badea, A. Dale e G. Johnson, "Automated Sewgmentation of the actively Stained Mouse Brain using Multi-spectral MR Microscopy," *NeuroImage,* vol. 39, nº 1, pp. 136-145, 2008.

[12] J. Nie e D. Shen, "Auromated Segmentation of Mouse Brain Images using Mulit-atlas Multi-ROI Deformation and Label Fusion," *Neuriinformatics,* vol. 11, nº 1, pp. 35-45, 2013.

[13] P. V. Hough, "Method and means for recognizing complex patterns," *U.S. Patent 3,069,654.,* 1962.

[14] M. Nixon e A. Alberto, Feature Extraction and Image processing, 3nd ed., Elsevier, 2012.

[15] R. C. Gonzalez e R. E. Woods, Digital Image Processing, 3rd ed., Pearson International Edition, 2008.

[16] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, nº 1, pp. 5-32, 2001.

[17] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," em *Graphics gems IV*, Academic Press Professional, 1994, pp. 474-485.

[18] M. Pietikäinen , "Local Binary Patterns," 2010. [Online]. Available: http://www.scholarpedia.org/article/Local_Binary_Patterns. [Acesso em 18 December 2016].

[19] N. Dalal e B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 05),* vol. 1, pp. 886-893, 2005.

[20] J. M. Bland e D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The lancet,* vol. 327, nº 8476, pp. 307-310, 1986.

[21] C. A. Tan, . S. Topper, C. W. Melver, J. Stein, . A. Reeder, K. Arndt e S. Das, "The first case of CDK5RAP2-related primary microcephaly in a non-consanguineous patient identified by next generation sequencing," *Brain and Development,* vol. 36, nº 4, pp. 351-355, 2014.