# Human Body Part Detection and External Injury Prediction Using Convolutional Neural Network

Aliasgar Haji[1], Riya Saraf[2], Dipti Pawade[3], Ashwini Dalvi[4], Irfan Siddavatam[5]
{mustafa.aliasgar@somaiya.edu[1],r.saraf@somaiya.edu[2],diptipawade@somaiya.edu[3]}

[1,2,3,4,5]Department of Information Technology, K. J. Somaiya College of Engineering, Vidyavihar, Mumbai, India

**Abstract.** Fake callers continue to disrupt emergency ambulance services in the state with its call center registering 7 percent fake calls every day on an average. This is a growing problem and it needs to be curbed at the earliest as it not only wastes the time of the operators but also keeps the line busy hence causing a delay in emergency services which may even result in the death of the victim. Hence, we propose a solution to validate whether the request for ambulance services is genuine or not. The main aim is to detect and identity from an image whether a human body part is present or not. Even if the image does not contain the entire human in any particular pose and only a part of his/her body is present. We also detect any visible external injury on the body part. This would be a proof of concept in the form of a machine learning model that can successfully detect feet, face, hands and any external injury present in the image provided to it.

**Keywords:** Body part detection, injury detection, Convolutional Neural Network (CNN).

## 1 Introduction

Hoax calls is a big issue for the state government's ambulance service providers [1-2]. One of the 15 calls received by the 108 emergency ambulance service is fake. Where the callers either make silent calls or use abusive language. More than 100 fake calls were received in only few months, by emergency services in Maharashtra ,alone. These callers, waste the call center operators' time. Some of them become abusive when the women refuse to talk and most of them cook up stories of road accidents. When officials reach the spot in the ambulances, they find nothing. These calls result in the wastage of resources because of which many times medical emergency services fails to reach to the genuine callers. It is essential to detect whether the accident is genuine or not. Hence, we propose a solution using image classification in which we detect various human body parts in the image and whether there is an external injury present or not. This validation system would provide as a screening for all the emergency calls made to such

ICASISET 2020, May 16-17, Chennai, India

DOI 10.4108/eai.16-5-2020.2303952

helplines and would help reduce them significantly. We chose the Multi-label Image classification method to solve this problem mainly because the image may contain more than one body part and some external visible injury. Hence, to propose a proof of concept we created a custom dataset of 765 human body parts images and trained are model on these images. In the end, the model can successfully detect what part of the human body is present in the image and also whether it is injured or not. Hence, this would be a viable solution to the problem of fake emergency reporting.

## 2 Literature Survey

Detecting and labelling of human body part is necessary to provide clues in analysing wound of the victim. Detecting body part separately is considered difficult due to inter-class variations exhibited. In [3] body part detection framework consist of Extreme Learning Machine classifiers ,likelihood scores for each segment detected as a body part and interframe voting strategy to identify the body parts in each frame for each video. In [4] human detection is done using parts based model which can be broken into four components: feature detection, part detection, spatial part scoring and contextual reasoning including non-maximal suppression on two datasets that is INIRA and PASCAL. Human body part recognition was performed in 4 major steps in [5] i.e reading and showing ,loading of a image and reading pixel by pixel then another process is to mask the image frame and find edges of the human body part and the last and final step is to formation of each recognized part after masking. Furthermore work performed in [6] human body part detection was done using AGEX interest point detector and this algorithm also build foundation for local detectors. The critical step of estimation of various human body parts was done in silhouettes detection in [7] which uses segmentation technique to detect salient body part and skin tone detection. Here estimation of body parts is done using 5 basic body key points and 7 sub key points. CNN is the most recent and well suited method in visual classification tasks. In [8] the problem of semantic segmentation is approached using CNN that is trained end-to-end to predict the class labels. Similarly the state-art-method that is CNN was used in [9] for extraction of features from image instead of traditional hand-designed image feature extraction methods. In the era of generic object detection and human body part detection RCNN is achieving great success using deep CNN. The [10] hybrid body part detector demonstrated the merits for partially occluded detection by integrating the scores of the individual part detectors based on the occlusion map. The highest merging score is the best configuration to evaluate the detection score of the human detector[13][14]. Here the detection of human body was performed and the result were analysed on testing performance and was found that missing rate for head was 23% and the missing rate for leg was found for 44% whereas the paper proposed gives missing rate for head and leg is 25% . From the literature survey it has been observed that most of the researcher has work on the body part detection in context with posture or position identification. We have not found any evidences about the body part detection along with injury prediction and this encouraged us to work in this area.

## 3 Methodology

The major challenges which in developing the proposed system was to have an appropriate database and design a model which can give two level classifications. In first level classification, model will predict whether the input image contain the body part or not and in second level envisage the presence of external presence of external injury on that body part.

### 3.1 Database Building and Pre-processing

Since there was no pre-compiled dataset available, we created custom dataset of 765 images in total. Images were scraped using Google's advanced search. The dataset contained images from four classes: Hand (275 images), Feet (241 images),Face (157 images), Injured (92 images). To maintain uniformity, all the considered input images are of .jpg type having dimensions 400*400. To maintain the diversity we have also considered images of different orientations. To organize the images in a structured format, stored in a folder which can be further used for training model. For training the model, a .csv file is maintined which has the training image names and their corresponding true label. The four columns belonging to four classes are the one-hot encoded columns. If an image contains feet its value will be 1, otherwise 0. The image can belong to 4 different classes.

### 3.2 Model Designing

For our model we have considered the following hyper parameters.

- Number of epochs: Initially, the model was trained with the number of epochs =10 and the accuracy achieved was 80%. Later after increasing the computation power we increased this number and saw a significant increase in the accuracy. The accuracy peaked at epoch =13 after which it started dropping.
- Shuffle Data: Random shuffling of data is necessary before training the model to avaoid the bias learning that may be caused because of a predefined data order feed to the model while training.
- Activation function: In our case, there can be more than one label for a single image but the probabilities need to be independent of each other. Thus, the sigmoid activation function is used which will predict the probability of each class independently.
- Loss Function: In order to improve the performance of the model, the loss should be minimized. This is acheived using the n-binary_crossentropy loss is used.

- Optimizer: The efficient Adam gradient descent optimization algorithm works well with the binary cross-entropy loss function and reduce the loss immensely.
- Dropout: Initially, dropout rate was set as 0.25 but was later reduced to 0.2 to increase the accuracy of the model.
- Hidden Layers: Total five hidden layers are added. Two convolution neural networks and two dense layers. Relu activation function is chosen. Relu is used in the hidden layers of the network to regularize the activation which should be zero or more than that and not negative.

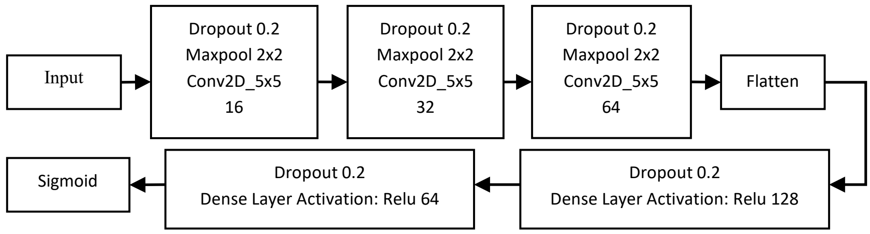



**Fig. 1** Model Architecture

Figure 1 depicts model architecture. The model first extract the features using convolutional layers and then sigmoid layer is used to predict the classes to which the image belongs. The model has an input of images of size 400x400 with RGB colour channel. The first layer of the model is a 2D convolutional layer of 16 filters with kernel size 5*5. Max pooling is then used to reduce the spatial dimensions of the output volume by 2 by setting pool size=2. Relu activation is applied along with dropout of 0.2 to help your network generalize and not overfit. The second layer is another conv2D layer with 32 filters and kernel size 5*5. Again, maxpooling with pool size 2 is used. The third layer is another 2D convolutional layer with 64 filters. We increase the number of filters so that the CNN learns more features from the images. The fourth and fifth layers are fully connected dense layers with 128 and 64 nodes respectively. The activation function used for each layer is Relu. The dropout of 0.2 is added to avoid overfitting of the above layers. The last layer is a dense layer with 4 nodes and sigmoid activation to perform learning from features extracted from the above CNN. The output of this layer are the prediction values themselves. We have used the n-binary cross-entropy loss function for training. The model is trained using the default learning rate of 0.01.


(a)

(b)

**Fig. 2** Model results with class label

For the experimental purpose, we have considered a total of 688 images for training and 77 images for testing. In the testing phase, the image is provided as an input to the model and get its prediction. The prediction is probability distributed among 4 classes and the largest is considered as the output.  We have displayed the top 2 classes for images having multiple labels. To calculate the testing accuracy we have tested our model on a database of 77 images. Few of the testing results are shown in figure 2. In figure 2(a), model has predicted a hand and it has injury while infigure 2 (b), the model has identified that image contains face but probability of injury is 0 that means no injuries.

## 4 Results and Discussion

The main objective of the proposed system is to detect which part of the human body is present in an image. Here, we have used multi-label image classification as more than body parts can be present in the image. We also wanted to detect whether there is any external injury present on any of the body parts.

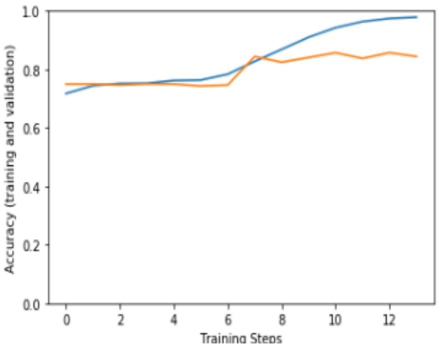



**Fig. 3** Model Accuracy Plot

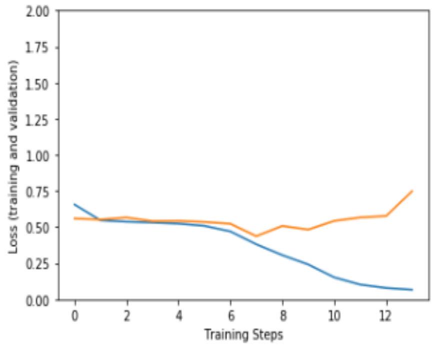



**Fig. 4** Model Loss Plot

We have trained are 7 layer CNN model on the training dataset. Initially, we got an accuracy of 81.3%. By increasing the number of epochs to 13, reducing the dropout rate and adding one dense layer of 64 filters we saw an increase in the accuracy. The training accuracy of our model is 97.42%, validation accuracy is 85.71% and the testing accuracy is 72.7%. Figure 3 shows the accuracy plot during the training phase. We see a dip in the validation accuracy in epoch 14 because the model starts overfitting. Similarly, the loss plots given in figure 4 shows a significant increase after epoch 13. For further analysis we plotted the confusion matrix shown in figure 5. Table 1 gives precision, recall and F1 score for each class. The average precision of the system is 0.741497, average recall is 0.727273, average F1 score is 0.72999 and Cohens kappa is 0.4601226993865031.

**Table 1.**Model Performance Parameter

| | Class 0 (Hand) | Class 1 (Feet) | Class 2 (Face) | Class 3 (Injured) |
|---|---|---|---|---|
| **Precision** | 0.57 | 0.76 | 0.76 | 1.0 |
| **Recall** | 0.63 | 0.75 | 0.8 | 0.67 |
| **F1 Score** | 0.6 | 0.76 | 0.78 | 0.8 |

From figure 6, we can see that due to less no. of injured images on which the model is trained there is high precision but low recall which results in returning very few false negetive results, but most of its predicted labels are correct when compared to the training labels. The feet class has the most number of true positives and a high recall relates to a low false-negative rate. In the ROC curve shown in figure 7, we got similar results that the feet class has the steepest curve and thus has the minimum the false positive rate. The area under the ROC curve is largest for the feet class hence it has the highest true-positive rate.

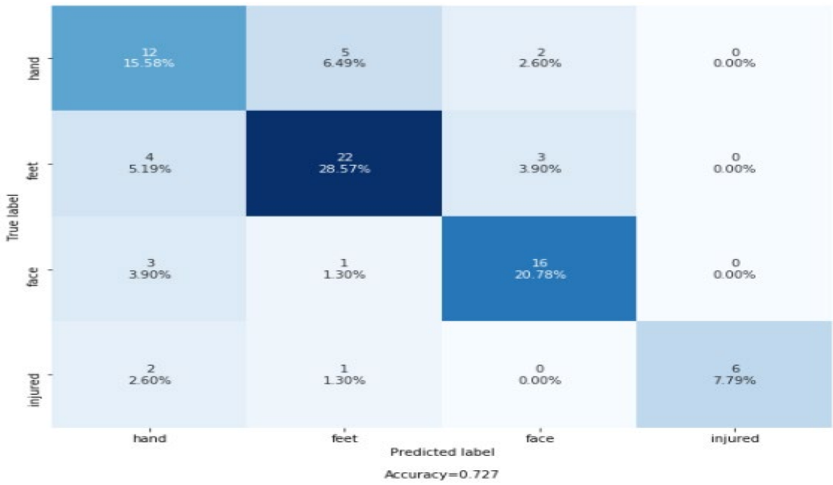
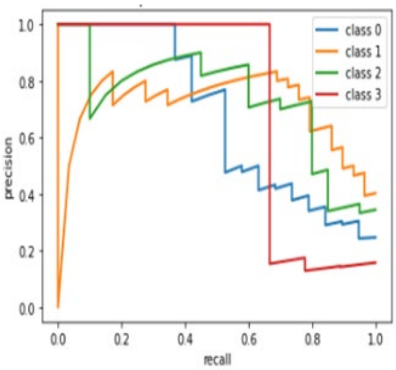



**Fig. 5**Confusion matrix



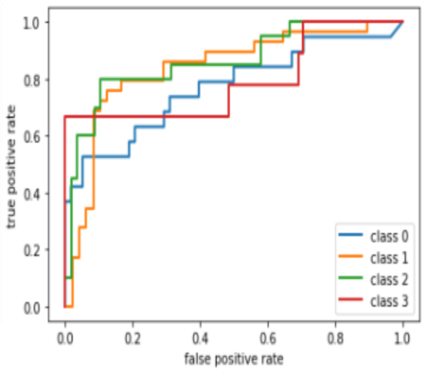
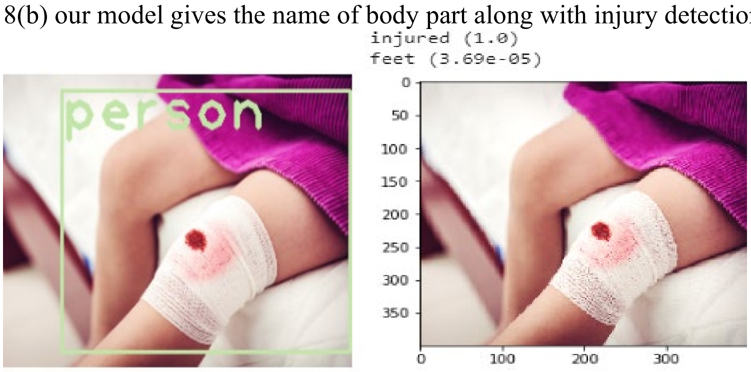




**Fig. 6** Recall vs Precision Plot **Fig. 7** False Positive Rate Vs True Positive Rate

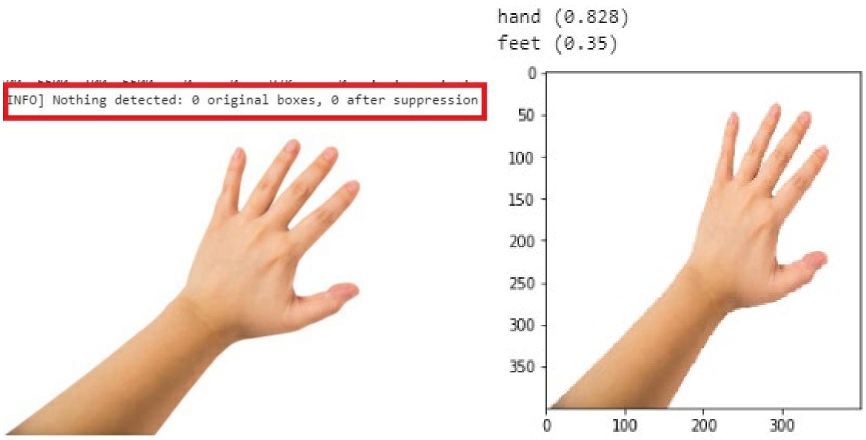
We have also compared our model with Yolo v3 model. YOLO v3 [11] uses a variant of Darknet, which originally has a 53 layer network trained on Imagenet. From the results we observed that YOLO v3 is unable to detect the different human body parts in the image. It can only detect a person in an image. Figure 8(a) shows the output of YOLO v3, where image is identified as a person while in figure 8(b) our model gives the name of body part along with injury detection.




(a) (b)

**Fig. 8** (a) Output of YOLO v3 (b) Output of proposed model

Again we have compared the results of our model with the other model [12] which has used a pre-trained HOG and Linear SVM detector based on the Dalal and Triggs method to detect people in images using INRIA Person Dataset (specifically, from the GRAZ-01 subset). As depicted in figure 9 (a) model is not able to detect anything, not even as person like in YOLO v3 model, while our model is able to identify it as a hand (figure 9 b).




(a) (b)

**Fig. 9** (a) Ouput of Model[12] (b) output of proposed model

## 5 Conclusion and Future Work

In this paper, we have discussed how we can use multi-label image classification for the detection of body parts and external injury using an image. Through the experimental results, we have observed the model accuracy as 72.7% and misclassification as 27.3% on the test data set. The accuracy of the system is 97.42% for the training set and 85.71% for the validation set. The model has a variance of 0.1171, a detection rate of 85% and miss rate of 15%. There is a huge scope of development is the project. The work can further be extended to do the classification of more body parts and accuracy can further be improved by increasing and diversifying. This will also generalize the model better for real-world detection.

## References

1. https://www.liverpoolecho.co.uk/news/tv/bbc-ambulance-fans-raging-over-16288397 Referred on: Aug 2019
2. https://timesofindia.indiatimes.com/city/jaipur/Fake-calls-by-ambulance-service-to-get-payments-RTI-report/articleshow/48288325.cms, Referred on: Aug 2019
3. M. Ramanathan, W. Yau and E. K. Teoh, "Human body part detection using likelihood score computations," 2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM), Orlando, FL, 2014, pp. 160-166.
4. Devi Parikh , C. Lawrence Zitnick . “Finding the Weakest Link in Person Detectors” . IEEE Conference on computer vision and pattern recognition ,(CVPR), 2011, Providence, RI,pp. 1425-1432.
5. Semih Yagcioglu ,Aykut Aras . “Human Body Part Recognition” . Eskişehir Osmangazi University, Eskisehir, TURKEY Senior Project Report, June 2008
6. Christian Plagemann , Varun Ganapathi , Daphne Koller , Sebastian Thrun . “Real-time

Identification and Localization of Body Parts from Depth Images” . Artificial Intelligence Laboratory, Stanford University, Stanford, CA 94305, USA.
7. A. Jalal, A. Nadeem and S. Bobasu, "Human Body Parts Estimation and Detection for Physical Sports Movements," 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE), Islamabad, Pakistan, 2019, pp. 104-109.
8. G. L. Oliveira, A. Valada, C. Bollen, W. Burgard and T. Brox, "Deep learning for human part discovery in images," 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, 2016, pp. 1634-1641.
9. D. T. Nguyen, H. G. Hong, K. W. Kim, K. R. Park, "Person recognition system based on a combination of body images from visible light and thermal cameras", Sensors, vol. 17, no. 3, pp. 605, 2017.
10. J. K. Kang, T. M. Hoang and K. R. Park, "Person Re-Identification Between Visible and Thermal Camera Images Based on Deep Residual CNN Using Single Input," in IEEE Access, vol. 7, pp. 57972-57984, 2019.
11. https://pjreddie.com/darknet/yolo/ Referred on: D3c 2019
12. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1.
13. Vinod Jagannath Kadam, Shivaji rao Manik rao Jadhav, K.Vijayakumar, “Breast Cancer Diagnosis Using Feature Ensemble Learning Based on Stacked Sparse Auto encoders and Soft max Regression”, Image & Signal Processing, springer, june 2019.
14. K. Vijayakumar , K. Pradeep Mohan Kumar ,Daniel Jesline, “Implementation of Software Agents and Advanced AoA for Disease Data Analysis”, journal of medical systems, Part of Springer Nature 2019.