



Non-invasive technique for real-time myocardial infarction detection using faster R-CNN

H. M. Mohan¹ · P. V. Rao² · H. C. Shivaraj Kumara³ · S. Manasa⁴

Received: 26 September 2020 / Revised: 22 December 2020 / Accepted: 14 April 2021 /

Published online: 12 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The medical history explores that Myocardial Infarction has been one of the leading factors of death in human beings since several decades globally. The researchers' key tasks are to emerge a novel real-time health vision-based monitoring system with added measurement features like high accuracy, robust, reliable, low-cost, low power with high data security. The main purpose of this research is to bestow an advanced non-invasive algorithmic approach for detecting the chest pain posture and fall posture based vital signs of Myocardial Infarction and analyzing the performance of a Faster Region-based Convolution Neural Network algorithm. This object detection computer vision technique is simulated for 3000 three-dimensional real-life indoor environment RGB color images for two datasets Nanyang Technological University Red Blue Green, and Depth dataset and private dataset-RMS trained datasets using TensorFlow object detection Application Programming Interface. The 3D RGB Images of NTU RGB database used for Vital Signs of Myocardial Infarction performance analysis is an improved approach. The simulation results have been compared with the existing works. The demonstrated results of ResNet-101 Faster RCNN showed the evaluated metric values: high mean precision and average recall value is a major contribution in this work.

Keywords Myocardial infarction · Vital signs · Mean average precision · Average recall · Faster region convolution neural network

✉ H. M. Mohan
mohanhm@gmail.com

P. V. Rao
raopachara@gmail.com

H. C. Shivaraj Kumara
shivrajepe@gmail.com

S. Manasa
manasa.s.athresha@gmail.com

Extended author information available on the last page of the article

1 Introduction

Many studies conducted worldwide show that cardio-related diseases have been the leading causes of instant and unpredictable deaths for several decades [46, 51]. Out of all the several reasons for the myocardial infarctions (or heart attacks) in human beings, the major risk factors are hypertension, stress, family history, smoking, dietary intake, and lack of physical activity [13, 45]. The recent reports demonstrate some common early heart attack symptoms that could be traced-out before hitting by severe heart attacks. Those are like: shortness of breath 47.9%, Jaw/Neck/Back pain 5.2%, Arm pain or Numbness 56.1%, lightheaded or dizziness 17.3%, sweating 13.8%, chest discomfort being the highest 84% [39]. The count increases globally for both sexes' life expectancy over 5.5 years with the age group of 66.5 to 72 from 2000 to 2016. The researchers claim that the chances of heart attacks cases increase in old age people [41]. According to the United Nations world population prospects survey, the percentage of the age group 65+ of the total population in most countries has exceeded 20% of the total population. Considering the world's population, the number of older people count could increase approximately up to 2 billion over the years 2000 to 2050 [46]. It is at all times a safe remedial solution for the old age people with cardiac diseases to remain at homes [45]. The work intends to identify the negligence ignored as mild pain. It could be fatal to ignore the signs of pain. It may not be the cases develop severe chest pain. The combined symptoms in a person could help in diagnosing the heart attacks immediately, but not just the severity of the chest pain. When a person suffers from a heart attack, instant medication within the first one hour may save a life [30].

The present work targets the old age people staying alone at home or working in internal workplaces while driving the vehicles. One of the major symptoms for sudden cardiac arrest is crushing chest pain [39, 41]. The dependency of older people gradually increases as per the survey results indicated several older people escalate over the years. The modern lifestyle has predominantly accountable for the loneliness of old aged people staying alone at homes. There are cases where they are often neglected, not taken care with timely assistance which could be fatal for their lives [44].

Computer vision is an integral part of medical imaging in today's era. In most of the real-time computer vision applications, the main surveillance is based on two concepts: object detection and localization [21, 26]. Here, at this juncture, a legacy from neural networks and its cognate learning advanced neural network algorithms and their developments in various applications will provide a path for the researchers to continue their development activities for object detection as a motivational objective. The exactness of the classification schemes of various images depends on the primary concepts of computer vision and its localization methods to detect the images.

The detection performance mainly focuses on the various design concepts of deep convolutional networks (ConvNets). This is to extract the enhanced features and classification of the images with inbuilt advanced convolutional architectures. Convolutional neural networks (CNN) normally target the fundamentals of image classification techniques that are applied in extracting the features from 2D and 3D image data. In 1998, CNNs experimented by Yann Le Cun in Bell Labs by combining two inbuilt approaches, namely convolutional neural networks and backpropagation to classify the handwritten digits [17].

Fog computing is commonly known as ‘fog networking or ‘fogging’ is a computing model of a decentralized network, where data and processing are placed in logical locations at the edge of the network rather than the cloud platform. Vaquero et al. [47] elaborated the basic

concepts of fog computing with an architecture built up with the number of heterogeneous ubiquitous and decentralized mobile nodes communication. It also simultaneous cooperation among them in addition to storage and processing tasks without third mediator intervention. Fog computing infrastructure offers a novel solution for data gathering, processing, applications and storage. Cloud computing platform for managing e-healthcare solutions presents challenges like latency, bandwidth, reliability, large volumes of data, context-awareness. Fog computing overcomes these disadvantages and emerges as an alternative solution to provide reliable infrastructure in health monitoring applications [9, 49].

The recent developments of add-on devices for Intelligent Video Surveillance IVS, such as sensors, embedded systems, wireless networks, and computer vision, have brought a new era to the researcher's knowledge to develop novel surveillance systems in health care systems. In health care camera-based monitoring applications like security accidents, crime diffusion, Traffic signaling and terrorist attacks, the main leading force is the increase in funding to target on their development of the intelligent monitoring systems and adapting the state of art techniques to make novel automated surveillance systems. These modules allow the IVS system to preprocess the acquired video feeds onboard and transmit only relevant information to the server(s) through the communication network [31].

Recently the state-of-art healthcare segment works were implemented using a fog platform. Iman Azimi, et al. proposed a real-time remote patient monitoring based on IoT approach for early detection of the patients' arrhythmia from ECG signal by extracting the features of the patients suffering from cardiovascular disease [3]. Vijayakumar et al. highlighted an intelligent system based on fog computing to determine mosquito-borne diseases. The fuzzy k -nearest neighbor classification algorithm is used to distinguish between infected and uninfected class [48].

The medical survey's main focus explores and adopts highly secured health monitoring data in object identification to help the doctors diagnose in medical image processing applications [15]. The Deep Learning (DL) method can be used for extracting the high-quality image features of the raw image from database to achieve high accuracy and high data security in image detection environments [5]. Thus, it overcomes the limitations of a traditional object detection method.

The main investigation of this proposed work is to bestow an advanced signaling alarm algorithmic approach for identification of vital signs of cardiac patients through 3D depth RGB color images (3000) mainly at remotely located areas to prevent from death. The monitoring system's performance is evaluated for two datasets Nanyang Technological University Red Blue Green and Depth dataset (NTU RGB + D) and private dataset-RMS. For the past two decades, the progressive rate of deep learning approaches in daily life is increasing due to the following key factors: i). Emerging trends adapted in training large scale datasets to achieve high data security and accuracy; ii). By incorporating high-performance parallel architectures such as GPU clusters; iii). Significant improvements in design architectural strategies and training image data schemes. Currently, deep learning techniques that are widely used in various applications such as: Sparse Auto encoder-based method, Recurrent Neural Network Long Short-Term Memory (RNN-LSTM), Restricted Boltzmann Machine (RBM) method, Generative Adversarial Network (GAN), etc. Generic object detection of images mainly concentrates on two factors: localization and classification. The two types of image object detection are: (i) Partition the image into generating region proposals and classify them into distinctive classes; (ii) Direct object detection with a unique framework is considered a regression or classification problem [21]. The revolutionary developments started in 2014 for object identification using the region-based convolutional network (R-CNN). R-CNN

classification further expanded into several methods such as Fast R-CNN, Faster R-CNN, and Mask R-CNN. Regression-based classification frameworks such as Multibox, Attention Net, Grid Convolution Neural Network (G-CNN), You Only Look Once-YOLO, Single Shot Detector-SSD, Deep Dense Shaped Descriptor-DDSD and Deeply Supervised Object Detector-DSOD [21]. This has been achieved with rapid advancements in smart Health Monitoring Systems (HMS) that made to overcome the drawbacks of traditional healthcare solutions. Yet there are some open issues and challenges faced by HMS domain [26, 44].

1.1 Angina pectoris and myocardial infarction

Nowadays, atherosclerosis is often mistaken to be a modern disease. Several references highlight that atherosclerosis with symptoms of pain in arm and breast was found in Ancient Egyptians from the mummification remains more than 3000 years back [8]. Around 250 years ago, an outstanding work some amount of Breast's disorders [12]. Heberden performed on 20 cases of angina pectoris open the doors for physicians and medical scientists to understand chest pain cases. Finally, the causes of angina pectoris were not well identified and classified for further investigation of the patients. Later, Edward Jenner's works, Burns and Parry contributed to developing dynamically with the association of angina pectoris and coronary arteries' disease. During the twentieth century, Mackenzie, Herrick, Keefer, and others understood that one of the origins of angina pectoris resulted in myocardial ischemia. It resulted from disease-causing cardiac hypertrophy or intraluminal Coronary artery disease and a balance between myocardial oxygen demand and coronary oxygen supply [11].

The majority of research studies explain the most common cause of Myocardial Infarction shown in Fig. 1. In the myocardium, the occlusion of the coronary artery originated by the rupture of an atherosclerotic plaque and later leads to coronary thrombosis (Clotting of blood). This might eventually activate the arterial spasm. Over a period of time in the wall of coronary arteries there are possibilities of formation of fibrous tissue and cholesterol in the form of plaques. The rupture of unstable plaques can lead to thrombus formation (Blood clot) that blocks the coronary artery. There is a high probability where a blockage of the right coronary artery may lead to inferior wall MI alone or with right ventricular or lateral wall infarction. The

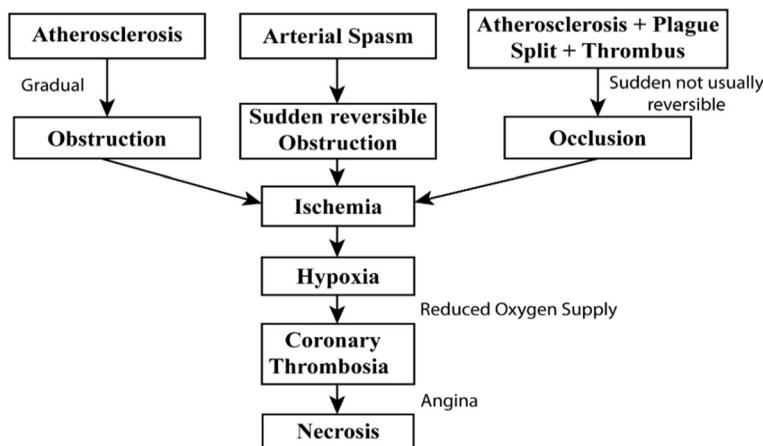


Fig. 1 Pathophysiology of myocardial infarction

implementation of blood flow to the heart persists, hampering the oxygen supply and demand to the myocardium that turns to the form ischemic cascade results to cellular damage. Finally, this can lead to necrosis and MI. Chest pain clinical investigations reveal multiple sources of causes which are categorized as in Fig. 2. They are mainly classified into three types: Cardiac ischaemic causes, cardiac non-ischaemic causes and non-cardiac causes [4]. Despite rapid advancements in cardiac health care analysis with advanced technology plays a major role, the distinction between non-cardiac chest pain and an acute Myocardial Infarction, AMI remains a dilemma [24].

The primary clinical investigations include electrocardiographic signal analysis, and the decision is being made based on the abnormality due to changes in signal characteristic. The most essential features tracked for probable MIs are prominent new ST segment elevation and new originated Q wave. Other powerful vital signs are the presence of a third heart sound, hypotension, and possibility of Chest pain diverging to both the right and left arm at the same time [29]. If the clinicians suspect MI, current practice is to put them under supervision in cardiac intense care unit and based on severity, the decision is being made whether to treat them with conservative management like thrombolysis or primary percutaneous transluminal coronary angioplasty.

The major contributions of this research work are:

- i) Creating of vital signs of heart attack detection with suitable annotation.
- ii) Finding a CNN backbone based Faster RCNN model for vital signs of heart attack detection.
- iii) Training a model to create vital signs detector and evaluating its performance.

This work is structured into the following sections as Section 2 figure out the correlated work with the remedial solutions. Section 3 provides a detailed description of the proposed methodology; Section 4 mainly highlights the experiments and results and Section 5 presents the analysis and discussion. Section 6 exposes the limitation in our work. Section 7 draws conclusions along with future work.

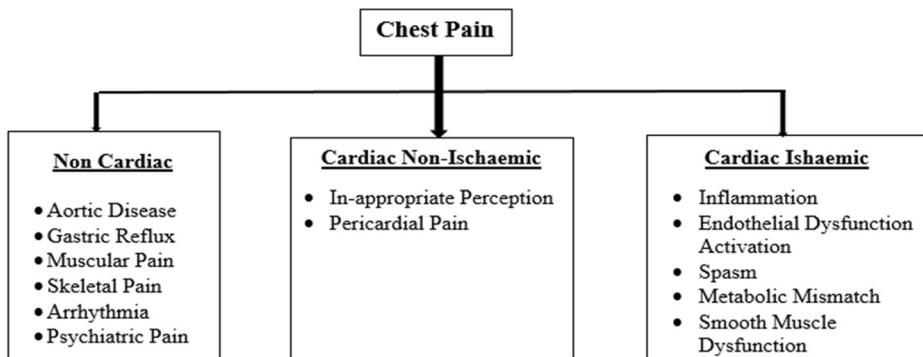


Fig. 2 Etiology of Angina Pectoris categorized as cardiac ischaemic, cardiac non-Ischaemic and non-cardiac

2 Related work

Artificial Intelligence has transformed the modern health care system and addresses complex medical challenges such as diagnosis, prediction and intervention. With the recent vast technological advancements in the medical domain, the concept of the future is to make it more of human independent and automated Artificial Intelligent (AI) driven [27, 33]. The medical survey's main focus explores and adopts a highly secured health data monitoring system in object identification to help the doctors diagnose. The major success in AI automated medical approach depends on working on vast datasets and building interpretable models. This, in turn, this helps in detecting complex diseases in real-time and makes precise medical decisions [38]. Many kinds of research have worked on the challenging health care issues, human activity recognition and human fall detection at different environments.

Health Monitoring Systems (HMS)'s success exists in collecting precise physiological, behavioral, and environmental data by understanding the patients' surroundings. It was processed accurately by suitable medical assistance [26]. Most researchers mainly considered the physical significance of older people problems. Depending upon the inputs reported in the survey, all these approaches are classified into visual and non-visual ways based upon the application. Wearable devices that use different kinds of sensors, like gyroscopes, microphones accelerometers, or combinations of all these can fall into the Non-visual approach category [2, 32, 40, 55]. Use of Kinect device, image and video-based recognition using cameras can be categorized as a visual approach [14, 19, 35, 42, 43]. Both approaches have their own advantages and disadvantages. Now-a-days, there are wide varieties of wearables available which can provide accurate solutions instantaneously. Wearable fall detectors like gyroscopes, accelerometers, or help buttons are used to detect falls.

The main drawback of using wearables is people should wear them all the time, which is practically impossible. Also, the sensors can suffer from misinterpretation of information which can be risky to lives. The heart rate trackers may require more R&D [2, 32]. Some of the traditional approaches to identify human activity is through human posture identification [40]. Computer vision approach offers a promising solution. There are some important factors considered for the video surveillance system. For reducing the artifacts and improving detection performance, some aspects are considered, including the type of high-end camera and its resolution, placement of the camera, and an efficient compression technique [43]. Considering standard imaging sensors, some approaches consider an input from the single-camera mounted on the ceiling, some mounted on the wall, multiple cameras placed to cover the entire room internally by capturing and understanding the 3D structure and objects accurately [14, 35]. Single-camera systems depend mainly on image space features extracted from silhouettes. Single-camera systems cannot capture 3D movable objects and extract features accurately to categorize the patients' fall condition.

Multi-camera systems extract the salient features such as velocity from three-dimensional objects that are modelled using back projecting multiple features. Identifying the cons of the camera-based approaches; several kinds of research have used Microsoft Kinect sensor to propose the fall detection by analyzing the depth images acquired [19, 42]. Deep learning algorithmic approaches have been tried in state-of-the-art human activity recognition, and this has been possible to adopt complex models on the same. Convolution Neural Network performs exceptionally well for solving computer vision problems. Many researchers explored various CNN based approaches for fall detection and other human activity recognition. The authors have proven that feeding human silhouette features from background-subtracted

RGBD picks up significant features used to evaluate the metric value for accuracy. The comparison shows the estimated value of extracted images is much better than the raw images. Although some excellent motivational works have been carried out for fall detection, only a little survey has been focused on chest pain vital signs detection. Few works provided justified solutions on early heart attack detection with color images using CNN technique [1].

One of the leading diseases to cause sudden death of human beings is a heart attack. Worldwide researchers have focused on the development of innovative and invasive approaches to detect heart attacks. However, very few researches have worked on the non-invasive heart attack detection methods. The author attempted to estimate two parameters with simulation results incorporating posture-based color images (1500), and the experimental results achieved better performance metrics for accuracy of 91.5% and sensitivity 92.85% [1]. It leaves a wide scope to the research in this area to save lives.

This proposed research work made a venture to boost non-invasive color image template matching algorithmic approach using a dataset of 3000 images. This could be an improved version of color image template matching technique with a high database that improves the overall performance in detecting the real-time heart attacks based on poster images. The risk factors can be analyzed by considering four possible posture positions- standing position, sitting position, partial fall and complete fall.

Recently there has been a groundbreaking work related to image processing field. The researchers provided three means clustering segmentation with salience region extraction solution to eliminate dehazing effect in certain area applications of natural scenes. To find an optimal solution using both visual and qualitative experiments to remove a haze of area such as sky/river [56]. The author developed a novel bat algorithm BABLU to achieve better performance quality with fast convergence. The experiments performed on Uni/multimodal with multi-dimensional problems and results are compared with other intelligent optimization algorithms [54]. The authors highlight the disadvantages of BBO in the multimodal register process in medical imaging registration and eliminated this using the best optimal solution Biogeography-based Optimization algorithm with Elite Learning (BBO-EL) using a hybrid full migration operator enhance matching characteristics of image registration [6].

In recent years, object proposal methods have proven to provide an unmatched performance [21]. They are categorized into two techniques, as follow: i) Sliding window technique. Edge Boxes, objectness in windows fall under this category. ii) Based on grouping superpixels. Multiscale Combinational Grouping, MCG, selective search, Cross Model Projection Classification, CMPC belongs to superpixels category. The selective search object detector methods; RCNN and fast RCNN [37] object identification approaches were inbuilt as an external module despite detectors dependency. The authors proposed a novel model of Faster RCNN based object detector to detect underwater seagrass leaf-based plant detection and annotation task of Halophile Oavalis to achieve high precision using inception V2 deep network [25]. The authors highlight the network architecture search on Neural Architecture Search Network, NASNET applied to build architecture to detect different objects using Faster RCNN for various applications such as Halophile Ovalis and other seagrass, pig feeding behavior in food with the metrics precision rate and recall rate, analyze MRI image features of regional lymph nodes in medical radiology applications and lymph node metastasis to diagnose rectal cancer for various databases in terms of diagnosis time of 20-s samples per case, and the vehicle traffic monitoring system at low altitude [53] [22] [52]. Impressed by the success of the Faster RCNN [34] algorithmic-based applications, the work proposes a new concept for a vital signs detection of myocardial infarction application using Faster RCNN

algorithm. There are inertial sensor based and environment based detection methods also exist and not suitable in noisy environments, RF-based methods are not preferable due to poor coverage area. Threshold-based algorithms mainly focused on wearable devices with less complexity and low recognition rate. To overcome these disadvantages, we adopted one of Non-threshold computer vision-based algorithm for better detection rate and fast computing speed. To achieve high accuracy with a low computational complexity of fall detection methods, Faster RCNN algorithm is preferred compared to other algorithms to improve the detection system's overall metrics performance.

3 Proposed methodology

3.1 Overview of the proposed model

Generally, CNNs can be used to classify the image identity within the bounded region or region of interest from the input image. The problem associated with this is identifying the spatial locations at different positions within the image and aspect ratios. This could cause a severe problem with an increased number of regions and computation time. The earlier existing RCNN and Fast RCNN algorithms for localizing the regions would increase computational time and slow down the processor performance and affect the network performance. Here the author introduces the following stages as follows: Faster RCNN, training process and detection method.

3.2 An outlook of R-CNN generation

The object detection models of Region-based CNN models are classified into, i). R-CNN for feature extraction and region proposals generation with different fixed-length feature vectors, ii). Fast R-CNN is an advanced version of the R-CNN model to detect the objects with one layer covering the entire image and the softmax classification layer to prevent a new prediction model generation. iii). Faster R-CNN- The drawbacks of RCNN (Generating region proposals in faster to bottleneck problems associated in Fast R-CNN) and the computation time of Fast RCNN is resolved by introducing a modified convolution network (Region Proposal Network, RPN) by replacing selective search algorithms. The Faster RCNN with inbuilt object detection and localization algorithm will be able to improve the performance metrics marginally includes accuracy and computation time as compared to RCNN and Fast RCNN models [21].

3.3 Introduction to faster RCNN

To eliminate the bottleneck problems of both RCNN and Fast RCNN, Ren et al. [10] describe a region proposal network with shared CNN layers in the network for object detection. Faster Region Convolution Neural Network performs two detection functions by localizing and categorizing the target image by incorporating, Region Proposed Networks and Faster RCNN detector. RPN generates rectangular region boxes. A faster RCNN detector proposes an additional network to predict reshaped proposals using pooling layers to grade the required image to fall within the presented region.

The algorithm is intended to reduce the computational burden and provide high performance of region proposal methods. It consists of two pipeline stages,

- i) Deep fully Convolution Neural Network: The first stage is for generating region proposals.
- ii) Fast RCNN detector: The second stage uses the proposed region.

Feature extraction: The color image is considered as input to a pre-trained convolution network. ConvNets are the backbone in implementing faster R-CNN object detection network architecture. In Faster R-CNN network, the high-quality image features are extracted by any one of the ConvNets such as Inception V2, ResNet-50, ResNet-101 in the first stage of object detection.

The main performance metric adopted in this process is Average Precision for Inception V2, ResNet-50, ResNet-101 and their performance is compared by training the dataset while extracting the features. ResNet is one of the solution providers for the bottleneck problem in deep CNNs to resolve the vanishing gradient issue due to jumping off one or more middle layers that leads to degrading the model's performance. ResNet model incorporates an identity matrix that enables to transfer the input information without any loss during the feed-forward path. The excellent performance metrics of Faster RCNNs can be obtained by increasing the architecture complexity in the object detection process. The region proposal network, RPN, is modelled as a sharable fully conventional neural network. RPN reduces the marginal cost and performs much better than other region proposal methods [34]. RPNs take an image of any dimensionality and produces the object proposals that measure a set of object class vs background. This RPN model shares computation with Fast RCNN model. The main impact of RPN on any faster RCNN is by adding additional layers to CNN. The features extracted by RPN are fed to box regression layers and classification layer to locate the region proposals and to obtain the target object in the box, respectively. The pooling layers of the network are mainly adapted to enhance the proposal box shapes before performing classification. The classification layer identifies the exact offset values of any bounding box. Thus the accuracy metric of the detection technique can be improved with Faster RCNN architecture.

Region proposals are created by sliding a small sliding window over the entire convolution feature map output layer. The $n \times n$ dimension of the convolutional feature map's spatial window is fed as an input to the selected network. Each sliding window is assigned with two-dimensional spatial coordinators. The smallest sliding window is referenced as a lower-dimensional feature. This feature is split into two fully connected layers, regression layer box and classifier layer box. Thus, the network implemented using a sliding window approach by providing the fully connected layers whose locations are shared over entire spatial local coordinates. In this mechanism, each location's maximum possible region proposals are detected at each sliding widow location. These maximum possible proposals are referenced as ' k '. The regression layer box provides $4k$ outputs related to k boxes' coordinates and classifier layer box indicates $2k$ scores that estimate the probability of object presence or not for each proposal. The ' k ' proposals measured with reference to k reference box are indicated as anchors. Figure 3 shows the architecture and region of proposal network.

Let, W =width, H =height of a convolution feature map with a yielding factor ' k ', that provides the total number of anchors (T) at each sliding position is given by (1). The architecture shown in Fig. 4 is a modified version of the existing one [7] in the detection process for video applications.

$$\mathbf{T} = \mathbf{W} \times \mathbf{H} \times \mathbf{k} \quad (1)$$

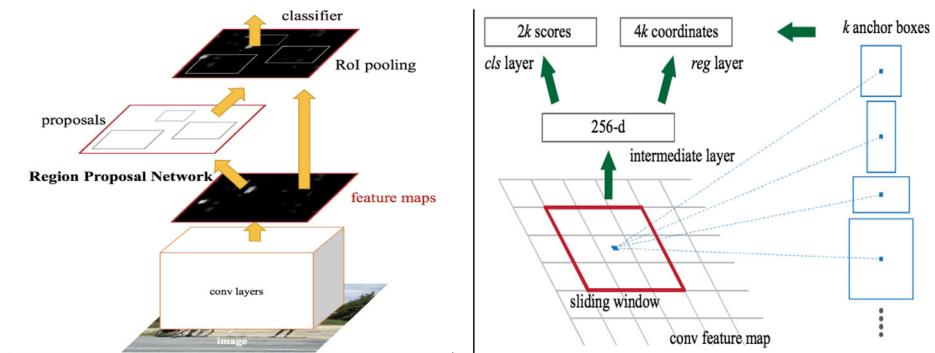


Fig. 3 Left: Architecture of Faster R-CNN Right: Region proposal Network RPN

3.4 The proposed method

This proposed work is sequentially organized into three stages: i) Input Stage, ii) Training stage, iii) Detection stage in real-time video. This work for myocardial infarction chest pain and fall posture detection uses Faster R-CNN architecture. The proposed vital heart attack detection framework using Faster RCNN algorithm is shown in Fig. 5.

The training images are prepared from the NTU 3D RGB database's video frames and for RMS dataset in the input stage. In the preprocessing technique, the images are downsized for improving the training speed. LabelImg software is used to create a region of interest for each image and convert every picture into XML format. During the training stage, the custom dataset is trained using Faster RCNN. Three ConvNet models pre-trained with COCO dataset: Inception V2 and ResNet-50 and ResNet-101(TensorFlow Zoo) are used for feature extraction and fine-tuning during the training of Faster RCNN model. Finally, in the detection stage, the camera detects the chest pain posture and fall detection from the video frames.

It indicates the successive stages incorporated in this Faster RCNN model for object detection. Figure 6 shows the flowchart of the proposed method.

The steps adopted during training and detection process with Faster RCNN as follows;

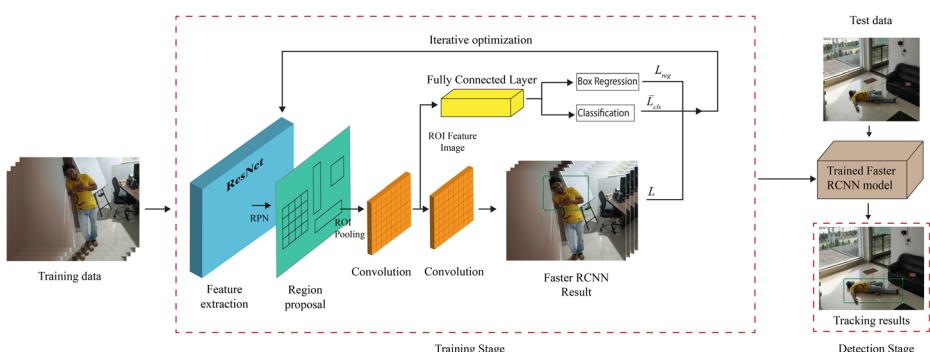


Fig. 4 The framework for training and detection on video-based monitoring of Vital Signs of MI based on Faster RCNN

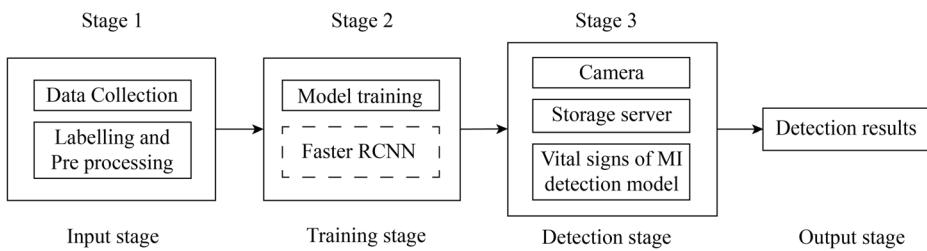


Fig. 5 Vital signs of heart attack detection framework with Faster RCNN

- Step 1: Set up of model environment using: TensorFlow library, Keras and TensorFlow Object detection API, OS-Windows 10, Open CV library and USB camera
- Step 2: Collect video frames from NTU 3D RGB database and create training images; and also, for RMS dataset. Labelmg software is used to create a region of interest for each image and convert every picture into XML format
- Step 3: Use 3 pre-trained models: Inception V2 and ResNet-50 and ResNet-101 (TensorFlow Zoo) and the custom data sets are trained using pre-trained models. An end-time solution is evaluated using the loss function and Mean average precision
- Step 4: Detection is carried out for the captured images by a camera and trained Faster RCNN Detector model to detect the vital sign postures of Myocardial Infarction.

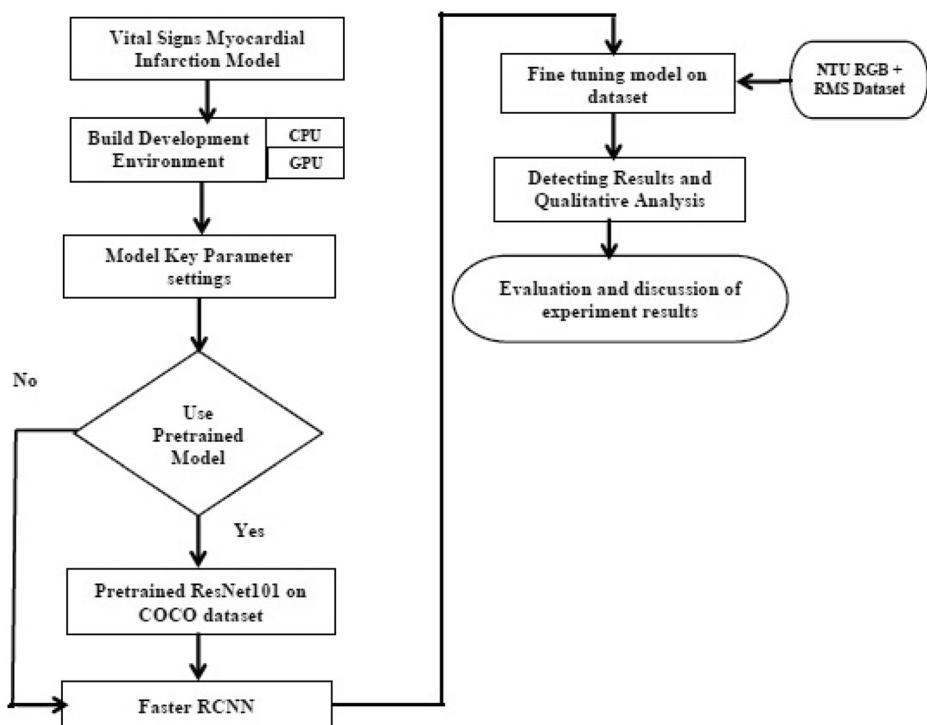


Fig. 6 Flow Chart of the Proposed Method- Vital Signs Myocardial Infarction Detection

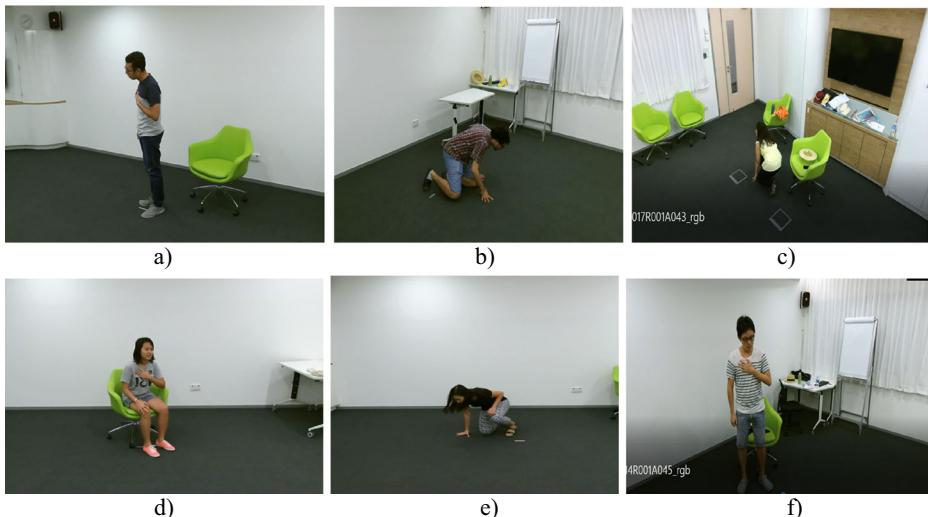


Fig. 7 Chest pain, Partial and complete fall posture images from NTU RGB + D dataset a) Chest pain standing posture b) Partial fall posture c) Partial fall posture d) Chest pain sitting posture e) Partial fall posture f) Chest pain standing posture

3.5 Situation identification: Different phases of vital signs detection

The scenarios of vital sign identification can be categorized into two phases:

3.5.1 First phase: Pain in the chest

At this first phase, chest pain occurrences in human beings are commonly noted in several ways. The best possible ways expressed for chest pain activity have been highlighted and taken into this database for further testing and diagnosis. The stabbing chest pain is sensibly captured in many ways, and some of them are expressed with physical behavior sensations such as: squeezed or pressed by a massive object and pain can extend from chest to the jaw, arms, back and to the neck. During the cardiac arrest, the person feels severe discomfort and often holds his chest. This posture can help us in capturing the first sign.

3.5.2 Second phase: Falling down

The person suffering from myocardial infarction tries to support a wall or any supporting structure nearby or falls instantly or eventually. The fall depends on the severity of the pain caused by human beings at the instant of time. Among the daily routine activities such as: sitting down, bending to take an object and other such activities identifying the fall is one of the most important problems. There are chances of misclassification which could raise a false alarm.

Fall detection principle and methods [28] are categorized based on the fall events as follows:

- **Pre-fall phase:** This event recognized is before the fall. Can be considered to daily life motions.

- **Critical phase:** This is during an extremely shortfall event. During the fall, the body is bent towards the ground, can be forward, backward or sideward.
- **Post-fall phase:** This is the most important phase to identify a person's fall and vital sign. The possible ways to occur this incident may fall into any one condition: falling on to the ground, motionless, struggling to walk or crawling with pain.
- **Recovery phase:** It is indicated absolutely by either the person stands up independently or else takes help from another person.

The characteristic postures help us capture one of the important vital signs of a heart attack: to discriminate the solution to the task using a computer vision algorithm. To track the signs mentioned above, each instance of the subject in the video frame is monitored. All the daily activities are being monitored that provide reliable information about a person to provide the individual situation at every instance of the time. The alarm is triggered while the person holding the chest and/or falls down for a predetermined time. This implies that each person probably caught up with the severe chest pain to be considered the vital signs indication to be monitored by us.

3.6 Data set collection and pre-processing

The datasets are collected from two sources based on benchmark standards; they are classified into two types as follows:

3.6.1 Action recognition dataset

Both NTU RGB + D dataset and NTU RGB + D 120 [36] [20] dataset together contain 1,14,480 video samples and 60 action classes totally. Every frame is captured by three Kinect V2 cameras simultaneously. Both datasets contain 3D skeletal data, IR videos, RGB videos, depth map sequences for each sample. In RGB videos, each frame has been captured with the 2D-pixel size of 1920×1080 ; and IR Video frame has a resolution of 512×424 . The actions of the dataset are categorized into three stages: a) Based on daily routine actions such as: drink water, stand up, put on a headphone, cutting nails, hand waving; b) Based on mutual actions such as shaking hands, hugging, take a photo; c) Based on medical conditions like- sneeze/cough, back pain, headache, chest pain, etc. The present work incorporates 3D RGB images of chest pain and partial fall postures from the dataset.

3.6.2 Private dataset- RMS

The authors have created a self-made dataset collected from the CCTV cameras. OnePlus 5 smart-phone camera of resolution 16MP set up at the indoor environment is labelled as RMS dataset. The original size of the image captured is of resolution 4608×3456 . The dataset contains different posture activities such as the poses, angles, scale, and lighting factors; and that are used to simulate the training dataset whose simulation results are compared with the simulated real-life heart attack fall scenarios. The chest pain images, fall with chest pain: partial and complete, are considered in this work.

Multiple scales and aspect ratio images can be trained using faster RCNN algorithm approach. The key idea is mainly to reduce the computational complexity so that the images can be scaled down to lower dimensionality. Image augmentation technique helps in better training the deep learning models improving the accuracy [16] [37]. In this work, the dataset is

collected in two ways one is standard NTU RGB + D dataset and NTU RGB + D 120 dataset, together with a huge number of samples. Another private data set RMS contains 1500 real-time heart attack condition simulated sampled images. The private RMS dataset images are captured under different lighting conditions. To enhance the processing speed and nullify the overfitting problem during the network's training period, these samples were downsampled to 1067×800 pixels. LabelImg annotation tool was employed to acquire the trained and tested sampled images. In this work, the sample date base assessed in three conditional posture stages: posture1-holding hand on chest (chest pain), posture2- partial fall with momentary changes, and posture3- complete fall on the floor Fig. 8.

Dataset 1 – NTU RGB + D Dataset – Collected 1500 images of sitting chest pain in sitting position and chest pain in standing position postures and partial fall postures with a resolution of 1240×600 . Figure 7 shows the pictures of NTU RGB + D.

Dataset 2 – Private Dataset - RMS – Collection of 1500 images of chest pain in standing position postures, partial fall and complete fall postures with a resolution of 1067×800 . Figure 8 shows pictures from Private RMS dataset.

A Region of Interest (ROI) of every image has to be selected, as shown in Fig. 9. A data labelling software is used to obtain the coordinates of ROI. The coordinates are later converted to XML format.

3.7 Implementing faster R-CNN algorithm to detect heart attack vital signs

The procedure is followed during the implementation of faster RCNN to detect vital signs identification of the patients. Raw Input images of three conditions of postures: chest pain posture, partial fall and complete fall are considered, and these inputs are to be preprocessed and later Faster RCNN algorithm used to simulate the detection of the vital signs of Myocardial Infarction. The steps adopted as follows:



Fig. 8 Chest pain, Partial fall & Complete fall posture images from RMS dataset a) Complete fall postures at home b) Complete fall postures at office c) Standing chest pain posture d) Chest pain standing posture e) Chest pain sitting posture f) Complete fall posture at the corridor

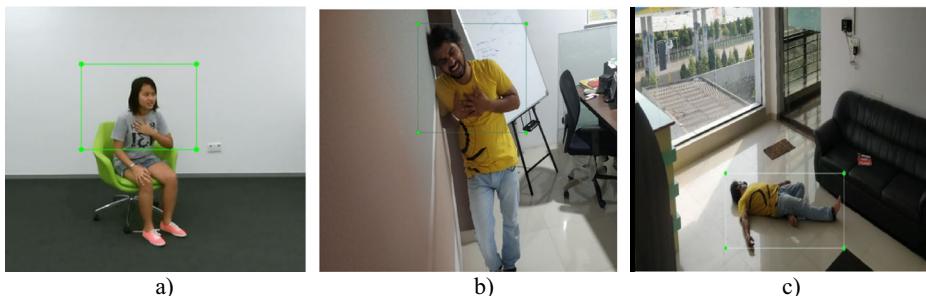


Fig. 9 Selecting the region of interest in the image. a) ROI selection for sitting chest pain posture. b) ROI selection for standing chest pain c) Region of interest for complete fall posture

- a. To get an input raw image from the dataset
- b. Procure high-resolution images into low-resolution images- cropped in preprocessing
- c. Set the Region of Interest ROI for all portioned images through annotation technique
- d. By extracting the features of raw sub-image by passing through ConvNet
- e. To evaluate losses, once the extracted images fed into Region Proposal Network (RPN) and scores have been measured for the object proposals - image sliding technique
- f. Apply ROI pooling layer to identify similarity in the object identification with predefined fixed-size feature maps.
- g. The object classification for the bounding boxes of the image can be performed more accurately and estimate performance metric values.
- h. Go back to c, until the number of steps completed

3.8 Training

The training process mainly focuses on initialization of the total convolution layers structure of Faster R-CNN. These shared layers were already trained by COCO dataset for ResNet101 model for classification. A deep learning framework, Object detection API uses a TensorFlow framework for the Faster RCNN Implementation in python. The standard Object detection API models have pertained with COCO dataset [18]. Training the model can be done using an optimization algorithm, stochastic gradient descent (SGD) and backpropagation technique. In Tensor flow object detection API, “image-centric” strategy from Fast RCNN has been used [34]. A fine-tuning technique has been used for training our model, retraining the existing model in this task. Fine-tuning was a prominent task for training this deep learning model for the supervised preprocessed data collected. We have used a learning rate of 0.0003 for our 3 k image dataset. The experiment setup has been kept under observation for the training time of 100 h for ResNet- 101 Faster RCNN model with the given configuration. The system configurations proposed in the current work are as follows: a) CPU: Intel(R) Core (TM) i7-7700 CPU @3.60GHz, b) Graphics card: Intel R HD Graphics 630, c) Memory: 12GB DDR4, d) Operating system: 64-bit Windows 10. High performance GPU can be incorporated to reduce the number of training hours for the present deep learning object detection model.

3.9 Performance evaluation metrics

In this proposed work, the authors have identified from the literature survey mainly six parameters during the evaluation process of Faster RCNN. In this architecture two losses that occur during the training of Region Proposal Network and classifier final stage. These losses can be dominated at RPN as objectness loss and localization loss and localization loss and classification loss as classifier stage. The four losses are estimated using loss function given by (2).

3.9.1 Loss function

The loss function for a single image $L\{p_i, t_i\}$ is explained and given by (2).

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (2)$$

Where,

I Index of an anchor in a mini-batch.

p_i ith anchor's predicted probability set an object.

p_i^* Ground truth label indicates anchor is +ve(−ve) for 1/(0).

t_i Vector with 4 local coordinates of the predicting bounding box.

t_i^* Ground truth box related to a + ve anchor.

L_{cls} Classification loss: the log loss due to two classes (object vs not object).

$p_i^* L_{reg}$ Regression loss is activated only for +ve anchors.

N_{cls}, N_{reg} Normalized weighted parameters for outputs of cls and reg layers and weighted by a balance parameter λ .

For bounding box regression, the localization points are expressed with 4 coordinates are as follows:

$$t_x = \frac{(x - x_a)}{w_a}, t_y = \frac{(y - y_a)}{h_a}, t_x^* = \frac{(x^* - x_a)}{w_a}, t_y^* = \frac{(y^* - y_a)}{h_a}$$

$$t_w = \log\left(\frac{w}{w_a}\right), t_w^* = \log\left(\frac{w^*}{w_a}\right)$$

$$t_h = \log\left(\frac{h}{h_a}\right), t_h^* = \log\left(\frac{h^*}{h_a}\right)$$

where, x, y, w and h indicate center coordinates, width and height of a Centre box respectively.

The variables x, x_a, x^* are assigned for predicting box, anchor box and ground truth box respectively (also for y, w, h). Thus, the design of anchors may help us in fixing the various sizes for the predict boxes. The classification problem of chest pain signs detection and fall detection demands how effectively faster RCNN performs. Micro averages of recall and precision are used as evaluation metrics. The Individual False negative (FN) cases and true positive cases (TP) of all WSIs are considered- these summation values are used to highlight the statistical preciseness.

3.9.2 Intersection over Union (IOU)

IOU is a measurement based on the Jaccard Index that indicates the overlap area between two bounding boxes. It is estimated using two boxes: ground truth bounding box B_p and predicted bounding box B_{gr} .

IOU estimates whether the detection is True Positive (TP) or True Negative (TN).

IOU is measured whose intersection area of the predicted bounding box and the ground truth bounding box divided by the area obtained by union operation. The IOU is given by (3),

$$\text{IOU} = \frac{\text{Area}(B_p \cap B_{gt})}{\text{Area}(B_p \cup B_{gt})} \quad (3)$$

A prediction is considered to be true positive (TP); if $\text{IOU} \geq \text{threshold}$. TP is a case of correct detection in which the bonding box sufficiently overlaps the ground truth.

A prediction is considered to be true positive (TP); if $\text{IOU} < \text{threshold}$. FP is a case of the wrong detection in which bounding box overlaps with the ground truth inadequately.

A ground truth that could not be detected can be termed as FN. True negative TN cases would represent a corrected misdetection.

3.9.3 Precision (P)

Precision (P) is the percentage of correct positive predictions. It is the performance factor of a model to identify only the required similar objects. This value is evaluated using (4),

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{TP}}{\text{Total number of ground truths}} \quad (4)$$

3.9.4 Recall (R)

Recall (R) is the percentage of true positive detected among all relevant ground truths. This value provides to estimate all the ground truth bounding boxes of a model (all the relevant cases). It is given by (5).

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{Number of predictions}} \quad (5)$$

3.9.5 Average precision (AP)

The average precision (AP) is the average value of precision reference to various recalls (j). Fix the value of each IOU and computes the mean of AP over N images will give us an indicator for the detection. AP metric is evaluated using (6).

$$AP = \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{j=1}^M \text{Precision}_i(\text{Recalls}_j) \quad (6)$$

Here, $Precision_i$ is a function of recalls_j. In this work, this metric plays a major role in association with a prediction score for each object and the model is evaluated as a different level of confidence.

3.9.6 F1 score

The F1 score realized by a constant value to indicate the success rate is expressed by two measurement metrics, precision and recall rate. In estimating the F1 score value both precision and recall rate will play equal priority to evaluate the performance of the framework, and F1 score is computed using (7)

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

4 Experiment and results

The Faster RCNN object detection API free open source is built on a TensorFlow framework. Keras, a user-friendly Application Programmable Interface (API) is being installed that increases the coding efficiency. Anaconda distribution is used with Python 3.7 environment. Keras, Tensor flow 1.14, NumPy compatible packages are installed via Anaconda3 for working of the model.

Figure 10 a), b), c), d) shows the experimental results for 3000 images collected from 2 databases NTU RGBD and Private RMS Dataset, and these results are analyzed using the tensor board for the predicted scores of the images. The Resnet-101 Faster RCNN algorithm effectively tested for 100 h to evaluate the test images of 500 images, and most of the test images cases it achieves 100% prediction rate. So, this classifier performance is much better than any other classifier during the evaluation process of the tested images.

Figure 11 exploits 3 possible classes based on the severity of the chest pain whose chest conditions with different postures are taken as the test case image analysis and performed the experiments for the model shown in this work ResNet-101 based Faster RCNN model. Here, the different postures were taken into consideration; chest pain posture, partial fall posture and complete fall posture to measure the predicted values for maximum accuracy, and this has been accepted as a motivational step in this work as a real-time application.

4.1 Object detection model training steps

-
- Set up Anaconda Environment
 - Get an input image from the dataset
 - Set the Region of Interest ROI and create class labels of all images
 - Convert Image to.xml data
 - Split dataset into train and test images in 80/20 ratio
 - Run the Training of Faster RCNN Algorithm using TensorFlow Object detection API
 - Visualize the simulation results on TensorBoard dashboard
-

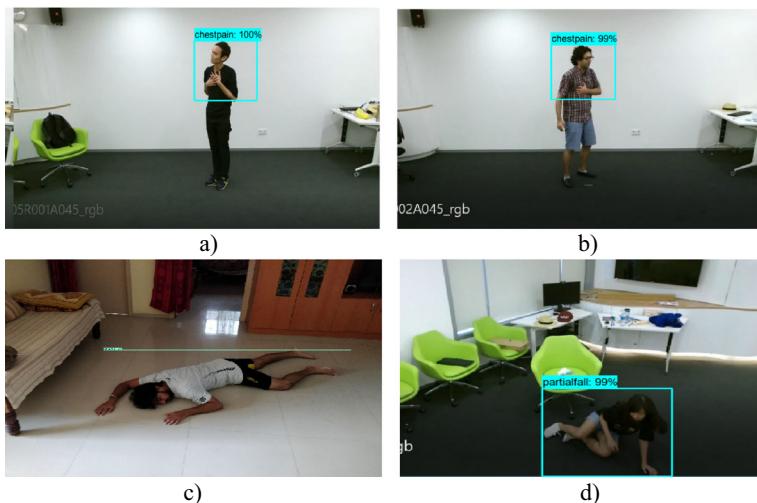


Fig. 10 Predicted simulated results of ResNet 101 based Faster R-CNN

4.2 Performance metrics evaluation

4.2.1 Precision evaluation

The graphical visualization results in regular intervals of checkpoints have been updated for all the parameters mentioned in this work. True Prediction: The 0.5 IOU ratio for each prediction at the training stage means detector predicts the overlapping area of an object bounding box with that of the ground truth box by at least 50%. When the localization is a matter for a

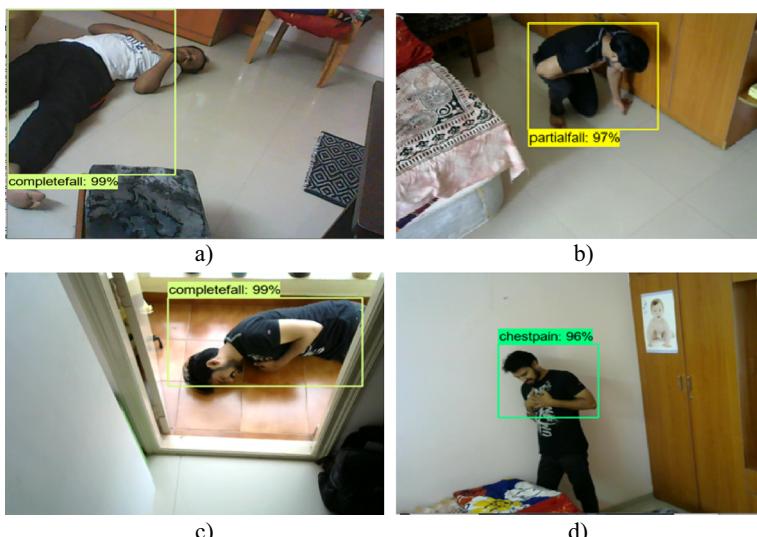


Fig. 11 Real-time detection of vital signs heart attack chest pain and fall image postures - score values. a) Complete fall posture in indoor environment b) Partial fall posture c) Complete fall in outdoor lighting environment d) Chest pain posture

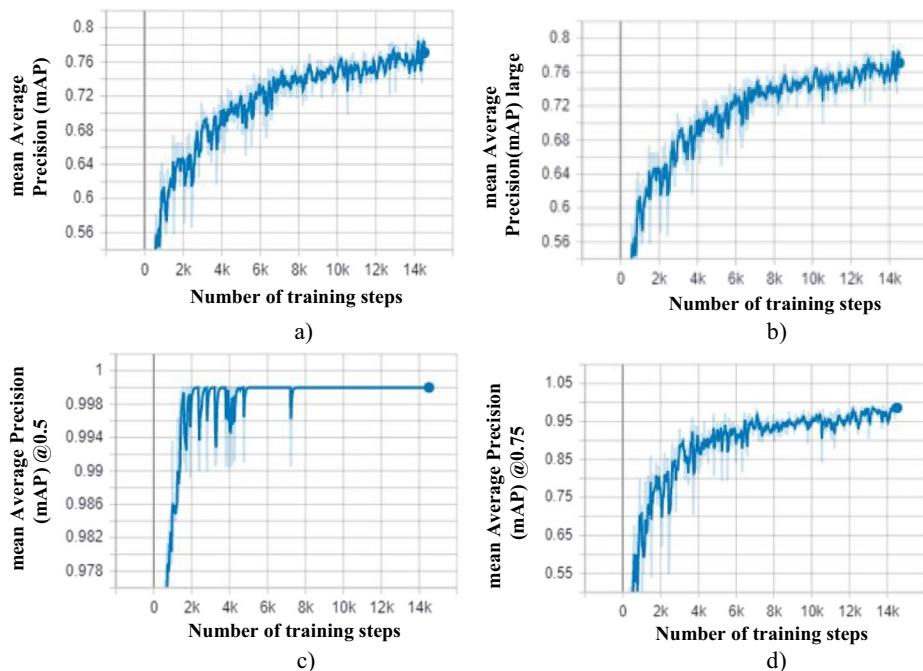


Fig. 12 Mean Average Precision values of ResNet 101 based Faster RCNN at different IOU. a) mAP values b) mAP (large) values c) mAP@0.5IOU d) mAP@0.75IOU

computer vision task, this ratio could be set higher. Regarding the standard representations of COCO evaluation metrics, both AP and mAP represent the unique identity in this proposed work. Similarly, the graph for mAP is plotted by computing by averaging over ten different IOUs whose values ranges from 0.5 to 0.95 with a step size of 0.05.

In this proposed Faster RCNN work, all the parametric measurements have been estimated for the IOU values at 0.5 and 0.75. Resnet-101 backbone based Faster RCNN model is trained in fourteen thousand steps for 100 h. The mAP values remain almost stationary after 14 K despite training for a longer period, as shown in Fig. 12, which is to be considered an optimal solution for the object detection model. If the model is trained for more steps, there might be a slight increase in the mAP value. The model results have been updated with 0.05 step size for the mAP value, mAP large value and mAP at 0.5 and 0.75@IOU in Table 1. The graphical window result shows the better prediction values which are obtained for mAP at 0.5IOU as compared to 0.75@IOU.

The simulated values obtained from the initial iteration value 560 to 14,000 for mAP metric are tabulated, starting from 560 to 14 thousand steps are tabulated in Table 2. The graphical

Table 1 Precision values of ResNet-101 based Faster RCNN at different IOU

Sr. No	Evaluation Parameter	mAP Value
1	Detection Boxes Precision (<i>mAP</i>)	0.785523866
2	Detection Boxes Precision/ <i>mAP</i> (large)	0.785523866
3	Detection Boxes Precision (<i>mAP</i>)@.50 <i>IOU</i>	1.0
4	Detection Boxes Precision (<i>mAP</i>)@.75 <i>IOU</i>	0.98235595

Table 2 The simulation results for mAP metric vs number of iterations with a step size

Metric/Steps	560	2 k	4 k	6 k	8 k	10 k	12 k	14 k
mAP	0.5565	0.6425	0.7165	0.7266	0.73	0.7386	0.7455	0.7855
mAP large	0.5565	0.6426	0.6691	0.6982	0.73	0.7445	0.7407	0.7855
mAP@0.5	0.9814	0.9918	0.9906	1.0	1.0	1.0	1.0	1.0
mAP@0.75	0.5494	0.6182	0.8133	0.8415	0.9521	0.9498	0.9519	0.9823

values indicate that the best optimal solution is obtained at @0.5IOU as compared to other results. The stationary values have obtained sharply at 4 k iterations @0.5IOU in object detection- bounding boxes process. The literature survey states that there is no optimum threshold selection in object detection. The solution for this is to identify an optimal threshold to be adjusted properly that mainly depends upon the object detection applications.

4.2.2 Recall evaluation

In this Faster RCNN object detection research work, three categories of postures are being used to detect chest pain, partial fall detection and complete fall detection. AR indicates the highest number of detections per image, averaged over categories and IoUs. The same is being elucidated by investigating the object bounding box areas. According to detection evaluation metrics, Common Objects in Context- COCO, mainly focus on predefined standard areas for objects such as i). Objects with smaller size $\leq 32^2$ pixels, ii). Object with medium size $> 32^2$ to $\leq 96^2$ and iii) Objects of large size $> 96^2$ pixels. The above standards mainly depend on the predefined area with the number of pixels existing in the segmentation mask. In this work both the datasets used during training consist mainly of large size segmentation masks.

4.2.3 Average recall (AR)

AR highlights the quality factor in object detection as the maximum number of detected results per image. The simulation experiment was performed to achieve the targeted outcomes in the quality measuring proposed in the object detection for vital signs of myocardial infarction using improved Faster CNN model. In order to attain this, two database sets have been incorporated to achieve a higher quality of test results. Figure 13 depicts the average recall values vs the different number of detections per image graphically. The experiments were set to evaluate the AR metric with three test conditions per image detection: 1, 10 and 100 detections per image. The same results of all three cases are tabulated in Table 3. The average recall value increases nonlinearly exponentially from the simulation results of average recall against the overlapped threshold values until the 10,000 iterations. After 10,000 iterations the slight increase in AR value (remains stationary). Finally, the AR values' performance evaluation highlights the similar functionality of the detector over a limited range of iterations.

Here the recall and precision parameters are interpreted with three possible conditions in object detection method, i) High recall – High precision, the highest number of objects are detected, ii) High Recall – Low precision, most detected objects are incorrect, (false positives). iii) Low recall - High precision (false negative) all predicted boxes are correct, and fault ground truth objects are many. The simulation results are mainly focused on the highest precision. The highest recall rate for the most true positive case and the same evaluated results are shown in Table 3 and Table 4, respectively. The simulated AR values are obtained by

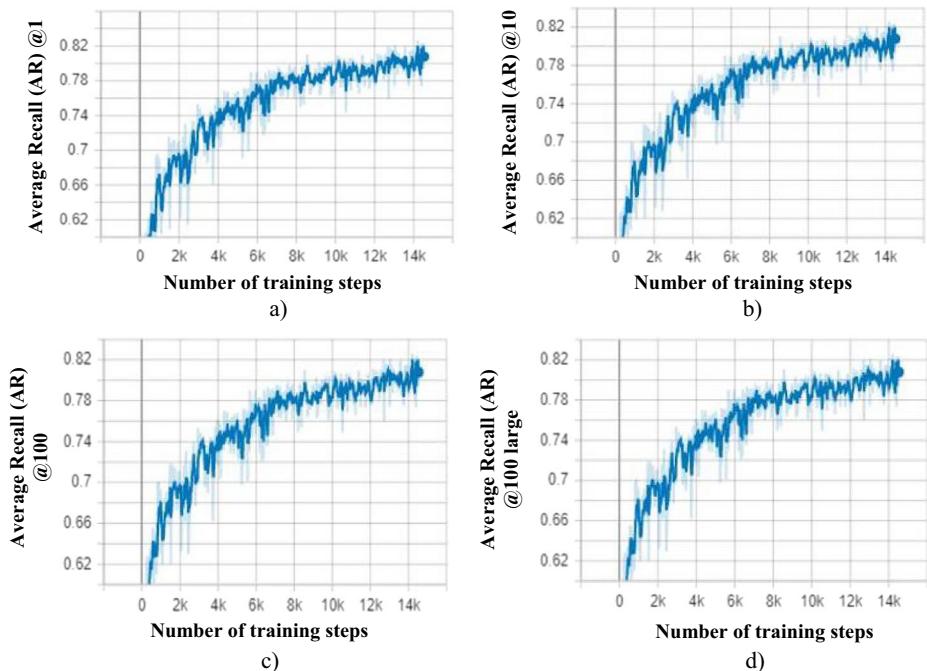


Fig. 13 Average recall values vs number of detections per image. a) AR@1 values b) AR@10 values c) AR@100 values d) AR@100(large) values

keeping the initial iteration value at 560, and it is evaluated until 14,000 iterations whose AR values are tabulated in Table 3.

4.2.4 F1 score evaluation

F1 score or F1 Measure or F measure used in binary classification schemes to indicate the measured value of a ‘test’s accuracy. The F1 score is judged with equal weights of both the precision and the recall factors of the test. Generally, F1-score evaluation is being adapted if both False Negatives and False Positives are crucial factors, else accuracy is adapted to measure under both True Positives and True Negatives (Table 5).

4.2.5 Loss function evaluation ($L\{p_i, t_i\}$)

The four losses happen at two different stages of a Faster RCNN network which are RPN as objectness loss and localization loss and classification loss as classifier stage. The four possible cases of loss functions were simulated using TensorFlow framework whose simulated results in four cases were represented in Figure 14(a), (b), (c), (d) and the overall loss function also represented in Figure 14(e). From the simulated results, it is observed that the loss function start decreasing at a lower number of iterations in 3 possible cases (a) (c) (d). As the number of iterations increases beyond 10,000 iterations, the loss functions get stabilized and intern, the overall loss function goes on decreasing; thus, the object detection model performance increases. Due to slight variation of the loss parameter RPN objectness loss the total loss function slowly reaching to stationary value in this work. Table 6 exploits the simulation

Table 3 Recall values of ResNet-101 based Faster RCNN at different IOU

Type	Evaluation Parameter	Value
a	Detection Boxes Recall (AR)@1	0.81178063
b	Detection Boxes Recall (AR)@10	0.81178063
c	Detection Boxes Recall (AR)@100	0.81178063
d	Detection Boxes Recall (AR)@100 (large)	0.81178063

results against the number of iterations for the loss function metric for the proposed Faster RCNN model.

The Faster RCNN architecture is modeled each one individually for backbone ConvNets Inception V2, ResNet-50 and ResNet-101. The performance of Faster RCNN model has been trained for 3000 RGB Images dataset. The prime metric of average value whose tested values with variable iteration steps are indicated for different ConvNets in Table 7. During training, the mean Average Precision evaluation starts at IOU threshold of 0.5 until up to 0.95 with a 0.05 step size. In the Faster R-CNN model, ResNet-101 shows much better performance than other models to achieve high detection rates on all classes.

4.3 Training time comparison

The training time of three ConvNets Faster RCNN model is shown in Table 8. In this work, while training the ResNet101 Faster RCNN object detection model, it was identified that one of the critical factors is training duration which takes longer period as compared to Inception V2 and ResNet 50 model. In this work the model was implemented with lower system configurations: Graphics card Intel R HD Graphics 630, Intel Core I7–7700 CPU. This can be eliminated by proposing higher system configurations with High-performance GPUs which can eliminate the longer execution times in object detection evaluation process.

The limitation of the simulation results in run time in the practical applications is 100 h. The annotation of training samples will lead to several days for getting the relevant results. In this paper, the whole dataset's annotation (3000 training images and 500 testing images) using the tool LabelImg more than ten days for two research fellows. It can lead the researchers much to rethink the methodologies to reduce the run time against the large database, which leads to further enhancing this work.

4.4 Results comparison with other work

Generally comparing the results of different papers working with deep learning algorithms has certain limitations even though authors use the identical database. Some of the reasons of differences are: i) Using subsets of custom-made data, ii) Performance measures, iii)

Table 4 Recall values vs Number of Iterations

Metric	560	2 k	4 k	6 k	8 k	10 k	12 k	14 k
Recall @1	0.6727	0.6846	0.7129	0.7717	0.778	0.7851	0.7768	0.8117
Recall @10	0.6727	0.6846	0.7129	0.7717	0.778	0.7851	0.7768	0.8117
Recall @100	0.6727	0.6846	0.7129	0.7717	0.778	0.7851	0.7768	0.8117

Table 5 Mean Average Precision, Recall and F1 Score values of the proposed Faster RCNN

Method	Backbone	mean Average Precision	Recall	F1 Score
The proposed Faster RCNN	Inception V2	71.1%	75.2%	73.09
	ResNet - 50	73.5%	78.5	75.91
	ResNet - 101	78.5%	81.7%	80.06

Evaluation methodologies followed, iv) Prediction task. As per the authors' extensive literature survey, some of the fall detection techniques have been evaluated using widespread object detection Faster RCNN algorithm. Table 9 shows the comparison of performance metrics of this proposed work with Xiang Wang [50] and Kun-Lin Lu [23].

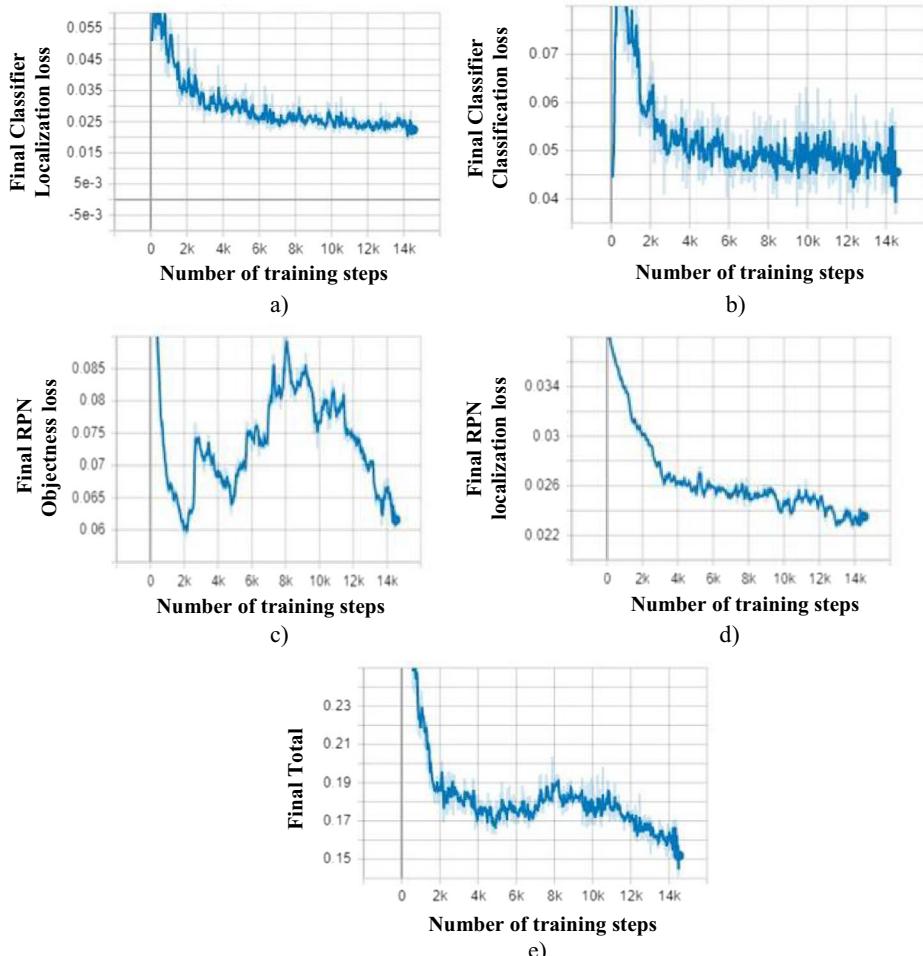


Fig. 14 Loss curves of training ResNet 101 based Faster R-CNN a) Final localization loss b) Final classifier classification loss c) Final RPN objectness loss d) Final RPN localization loss e) Total loss function

Table 6 Various loss factors measured by ResNet 101 based Faster R-CNN model

Type	Loss Type	Loss value
A	Final classifier localization loss	0.02201158
B	Final classifier classification loss	0.04522482
C	Final RPN objectness loss	0.073151
D	Final RPN localization loss	0.024450922
E	Final Total loss	0.16483831

5 Analysis and discussion

Object recognition has evolved as a challenging and major focus area for researchers in the computer vision domain. In this paper, the authors attempted to enhance the proprieties and evaluate the Faster R-CNN is a state-of-the-art deep learning algorithm, for the vital signs of heart attacks. The significance in object detection for the severity of chest pain based vital signs detection using binary classification. The prime importance of developing this modified Faster R-CNN technique for posture chest pain detection with predefined three possible posture-based analysis from the databases and this development allows the doctors to easily diagnose the patient condition to perform the task effectively efficiently in hospitals. Thus, a posture-based database can help the doctors in future investigations. Myocardial Infarction provides maximum reliable information about the patients with chest pain as real-time monitoring application. The in-house environment with different possible phase and fall detection postures is a major contribution to this work.

Faster R-CNN utilizes the pyramids of reference boxes, which serve as references at multiple scales and aspect ratio and are called “anchor” boxes by the authors, in an innovative way. Further, since a single CNN was used to carry out chest pain region proposals and classification, only one CNN needed to be trained. We obtained region proposals with almost zero computational cost. The prime factor selective search is implemented using a single convolutional neural network-CNN as Region localization rather than a computationally expensive algorithm (e.g. R-CNN and fast R-CNN), and this can achieve the highest accuracy for the model with assigning k different object rectangular shaped boxes [tall, wide, and large box]. To evaluate the object detection algorithm, an experimental work was performed to find the best test- train configuration ratio and to obtain the best training model by yielding minimum loss criteria. This work pinpoints the localization errors that may be caused in a Faster R-CNN for chest pain posture-based detector model due to the IOU scaling values of AR performance parameter, and this has to be taken into consideration to enhance the performance of Faster RCNN model in future. Most of the works from benchmark NTU RGU + D dataset considered video images obtained by depth cameras as the input for their algorithmic approaches. This proposed research work considers the input images from RGB videos of NTU RGB + D and private RMS dataset captured from high-resolution cameras.

Table 7 Measurement of AP with different ConvNets – Faster RCNN

Method	Backbone	AP %	AP ₅₀ %	AP ₇₅ %
Faster RCNN	Inception V2	71.1%	98.3%	88.8%
	Resnet - 50	73.5%	100%	97.2%
	Resnet - 101	78.5%	100%	98.2%

Table 8 Measurement of training time, number of steps: Types of ConvNets - Faster RCNN

Object Detection Architecture	DCNN	Training Time (Hrs)	Number of Steps
Faster RCNN	Inception V2	35	20 k
	Resnet-50	60	8 k
	Resnet-101	100	14 k

6 Limitations

Here in this work, few gaps were found, and those are discussed below as follows: The current proposed work is implemented using two sets of databases: one public database and another private database in detecting the vital signs of chest pain from the 3 posture models. The Faster R-CNN object detection model for chest pain poster model's main limitation is limited available databases with different posture conditional images. Secondly, this work can be enhanced using data augmentation technique to improve the quality of images identification. Third point, AR and Precision metrics can be analyzed for the different IOU scaled values; instead, the attempts made for the fixed values found in the survey. Finally, one more factor can be taken into the discussion is an increase in training steps, and the reports are analyzed during the object detection methodology.

7 Conclusion and future work

Automating pain detection and estimating the pain intensity level via suitable pain management strategies can emerge as a lifesaver in medical health informatics. The artificial intelligence approach adopted in this research plays a significant role in researchers and medical professionals working with pain management practices. The authors developed a comprehensive solution of fog computing network real-time chest pain posture-based pain detector that guarantees as a pain estimator and detection solution under myocardial infarction emergency condition to save lives. The conclusions are made based on the simulation results obtained for myocardial infarction vital signs detection for two databases. Here, the authors proposed two data sets to accurately perform one simulation for real-time vital signs-person detection using public and private datasets using TensorFlow object detection API. In this proposed object detection model, we utilize Faster RCNN by using transfer learning with pre-trained network architectures: Inception V2, ResNet-50 and ResNet-101. The deep learning model was evaluated for five performance metrics to perform better posture detection of chest pain, partial fall, and complete fall. The experimental results demonstrate that deeper and heavy model ResNet-101 based Faster R-CNN can achieve the highest mean Average precision (78.5%) and Recall (81.1%) values. The other two less deeper models Inception V2 and ResNet-50 comparatively achieve lesser performance. The object

Table 9 Measurement of mAP metric of this proposed work

Paper	Model	Dataset	mean Average Precision
Xiang Wang [55]	Faster RCNN with Resnet - 101	Private open dataset	76.4%
Kun-Lin Lu [56]	Faster RCNN with NASNet	Private open dataset	75.8%
Our work	Faster RCNN Resnet - 101	Custom RMS database and NTU RGB	78.5%

detection vital signs of Myocardial Infarction model developed were able to generalize across different environmental conditions. People also differentiate between Vital signs considered and other everyday human movements. The results presented indicate that Faster R-CNN could be utilized in practical settings as a method of early signs detection of myocardial infarction. The provision is made with hardware development based development-based alarm signal indicator for every instance of fall detection identification, and the same is forwarded to near hospitals or caretakers of the person to get better assistance immediately. The temporal information of video frames and Time and Space complexity may be considered in the future to evaluate our object detection system's performance.

Acknowledgements The authors would like to express sincere thanks to the organization Digital Shark Technology, Bangalore, for providing hardware resources during the implementation of this work. We thank Dr. Steve Ling, School of Biomedical Engineering, University of Technology, Sydney, Dr. R Srinivasan Ex. Dean R&D, RNSIT, Bangalore and Dr. Rifai Chai, School of Software and Electrical Engineering, Swinburne University of Technology for their encouragement and suggestions during this research work.

Declarations

Declaration of conflict of interest The authors declare that there is no conflict of interest.

References

1. Albaracín GR, Chaves MA, Caballero AF (2019) Heart Attack Detection in Colour Images Using Convolutional Neural Networks. *Applied Science* 9:5065–5074
2. Amrawy MN (2015) Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial. *Health Informatics Research* 21(4):315–320
3. Azimi I, Anzampour A (2016) Medical warning system based on internet of things using fog computing. International Workshop on Big Data and Information Security
4. Balla C, Pavasini R, Ferrari R (2018) Treatment of angina: where are we? *Cardiology* 140:52–67
5. Bizopoulos P, Koutsouris D (2019) Deep learning in cardiology. *IEEE Rev Biomed Eng* 12:168–193
6. Chen Y, He F, Li H (2020) A full migration BBO algorithm with enhanced population quality bounds for multimodal biomedical image registration
7. Chen T, Jiang Y, Wang J (2020) Maintenance personnel detection and analysis using mask-RCNN optimization on power grid monitoring video. *Neural Process Lett* 51:1599–1610
8. David AR, Kershaw A, Heagarty (2010) Atherosclerosis and diet in ancient Egypt. *Lancet* 375:718–719
9. Fog Computing in Healthcare (n.d.) A Review and Discussion. IEEE, Vol 5:9206–9222
10. Girshick R (2015) Fast R-CNN. International Conference on Computer Vision, pp 1–9, [arXiv:1504.08083 \[cs.CV\]](https://arxiv.org/abs/1504.08083)
11. Gorlin R (1996) Pathophysiology of cardiac pain. *Circulation* 32:138–148
12. Heberden W (1987) Some account of the disorders of the breast. *Clin Cardiol* 10(3):211–213
13. Hildreth CJ, Burke AE, Glass RM (2003) Risk factors for heart disease. *The Journal of American Medical Association*, 290(7)
14. Hur T, Bang J, Huynh-The T, Lee J, Kim J-I (2018) A Novel Signal-Encoding Technique for CNN-Based Human Activity Recognition. *Sensors* 18:3910–3929
15. Johnson KW, Soto JT, Glicksberg B (2018) Artificial Intelligence in Cardiology. *J Am Coll Cardiol* 71(23): 2668–2679
16. Lajczyk M, Grochowski (2018) Data augmentation for improving deep learning in image classification problem. International Interdisciplinary PhD Workshop, pp. 117
17. Lecun Y, Bottou L, Bengio Y (1998) Gradient-based learning applied to document recognition *Proc. of the IEEE* 86(11):2278–2324
18. Lin Z, Doll (2015) Microsoft COCO: common objects in context. *Computer Vision, ECCV* , pp. 740–755

19. Lin HY, Hsueh YL, and Lie WN (2016) Abnormal event detection using Microsoft Kinect in a smart home. International Computer Symposium,
20. Liu J, Shahroudy A, Perez M (2019) NTU RGB+D 120: a large-scale benchmark for 3D human activity understanding. *IEEE Trans Pattern Anal Mach Intell*, 2019.
21. Liu L, Ouyang W, Wang X (2020) Deep learning for generic object detection: a survey. *Int J Comput Vis* 128:261–318
22. Lu Y (2018) Identification of metastatic lymph nodes in MR imaging with faster region-based convolutional neural networks. *American Association for Cancer Research* 78(17):5135–5143
23. Lu K-L, Chu ET-H (2018) An image-based fall detection system for the elderly. *Appl Sci* 8(10):1995–2026
24. Malik MA, Khan S, Safdar S (2013) Chest pain as a presenting complaint in patients with acute myocardial infarction. *Pakistan Journal of Medical Sciences* 29(2):565–568
25. Moniruzzaman MD, Islam SMS, Lavery P (2020) Faster R-CNN Based Deep Learning for Seagrass Detection from Underwater Digital Images Digital Image Computing: Techniques and Applications (DICTA), IEEE
26. Mshali H, Lemlouma T, Molone M (2018) A Survey on Health Monitoring Systems for Health Smart Homes *International Journal of Industrial Ergonomics* 66:26–58
27. Mshali H, Lemlouma T, Moloney M, Magoni D (2018) A survey on health monitoring systems for health smart homes. *International Journal of Industrial Ergonomics* 66:26–56
28. Noury N, Fleury A, Rumeau P, Bourke AK, Laighin O (2007) fall detection – principles and methods pp 1663–1666
29. Panju AA, Hemmelgarn BR, Guyatt GH (1998) Is this patient having a myocardial infarction. *JAMA* 280: 1256–1263
30. Patel A (2018) Awareness of Heart Attack Signs and Symptoms and Calling 9–1–1 Among U.S. Adults. *Journal of the American college of cardiology*, 71(7)
31. Prati A, Shan C (2019) Sensors, vision and networks: from video surveillance to activity recognition and health monitoring. *Journal of Ambient Intelligence and Smart Environments* 11:5–22
32. Profis S (n.d.) Do wristband heart trackers actually work? A checkup. CNET, www.cnet.com/news/how-accurate-are-wristband-heart-rate-monitors.
33. Rashidi P, Mihalidis A (2013) A survey on ambient-assisted living tools for older adults. *IEEE Journal of biomedical and health informatics* 17(3):579–590
34. Ren S, He K, Girshick R (2016) Faster R-CNN: towards real-time object detection with region proposal networks. *Computer Vision and Pattern Recognition*
35. Sahoo SP, Ari S (2019) On an algorithm for human action recognition. *Expert Syst Appl* 115:524–534
36. Shahroudy A, Liu J, Ng T-T, Wang G (2016) NTU RGB+D: a large scale dataset for 3D human activity analysis. *IEEE Conference on Computer Vision and Pattern Recognition*, pp:1010–1019. <https://doi.org/10.1109/CVPR.2016.115>
37. Shorten C, Khoshgoftaar T (2019) A survey on image data augmentation for deep learning. *Journal of big data*, 6(60)
38. Siegersma KR, Leiner C (2019) Artificial intelligence in cardiovascular imaging: state of the art and implications for the imaging cardiologist. *Neth Hear J* 27:403–413
39. Smith KL, Cameron P, Meyer A (2002) Knowledge of heart attack symptoms in a community survey of Victoria. *Emerg Med* 14:255–260
40. Sokolovaa M, Cuerdab JS, Castillob JC (2013) A fuzzy model for human fall detection in infrared video. *Journal of Intelligent & Fuzzy Systems* 24:215–228
41. Stern S, Behar S, Gottlieb S (2003) Circulation, Aging and Diseases of the Heart circulation, 108(14)
42. Stone E, Skubic M (2015) Fall detection in homes of older adults using the Microsoft Kinect. *Journal of Biomedical and Health Informatics, IEEE* 19(1):290–301
43. Sung J, Ponce C, Selman B and Saxena A (2011) Human activity detection from RGBD images. *Association for the Advancement of Artificial Intelligence*, 47–55
44. Tang D, Yusuf B, Botzheim J (2015) A novel multimodal communication framework using robot partner for aging population. *Expert Syst Appl* 42:4540–4555
45. Theodorou P, Hatzis G, Nikolaos P (2017) Socioeconomic status and risk factors for cardiovascular disease: impact of dietary mediators. *Hell J Cardiol* 58(1):32–42
46. United Nations, Population Prospects (2019), www.un.org/en/development/desa/population/publications/trends/population_projections.asp.
47. Vaquero LM, Rodero-Merino L (2014) Finding your way in the fog: towards a comprehensive definition of fog computing. *ACM SIGCOMM Computer Communication Review*, 44(5)
48. Vijayakumar M (2018) Fog computing-based intelligent healthcare system for the detection and prevention of mosquito-borne diseases computers in human behavior

49. Vilela PH, Rodrigues J (2020) Looking at Fog Computing for E-Health through the Lens of Deployment Challenges and Applications Sensors, 20
50. Wang X, Jia K (2020) human fall detection algorithm based on YOLOv3. IEEE 5th international conference on image, Vision and Computing,
51. World Health Organization, World Health Statistics Overview (2019) www.who.int/gho/publications/world_health_statistics/2019/en.
52. Xu Y, Yu G, Wang Y, Wu X (2017) Car Detection from Low-Altitude UAV Imagery with the Faster R-CNN" Journal of Advanced Transportation, Vol 2017, Article ID 2823617
53. Yang Q, Xiao D, Lin S (2019) Feeding behavior recognition for group-housed pigs with the faster R-CNN. Comput Electron Agric 155:453–460
54. Yong J-s, He F, Li H-r (2019) A novel bat algorithm based on cross boundary learning and uniform explosion strategy. Applied Mathematics-A Journal of Chinese Universities 34(4):482–504
55. Yu G (2017) Ensembles of deep LSTM learners for activity recognition using Wearables proceedings of the ACM on interactive, Mobile, Wearable and Ubiquitous Technologies, 1(11)
56. Zhang J, He F, Chen Y (2020) A new haze removal approach for sky/river alike scenes based on external and internal clues. Multimed Tools Appl 79(20)

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

Affiliations

H. M. Mohan¹ • P. V. Rao² • H. C. Shivaraj Kumara³ • S. Manasa⁴

- ¹ Department ECE, T John Institute of Technology, Visvesvaraya Technological University, Belagavi, India
² Department of ECE, Vignana Bharathi Institute of Technology, Hyderabad, India
³ Skillmine Technolgy Consulting Private limited, Bangalore, India
⁴ Department EEE, Dayananda Sagar Academy, Visvesvaraya Technological University, Belagavi, India