Detection and Altering of Aggressive Behaviour of Mentally Unstable People Using CNN AND RSSI

Balaji. S. R.[1] , E. M. Ramya[*2], S. Padmashree[3], A.Nivethitha[4], N.Rachana[5]
[1]Associate Professor, Panimalar Engineering College, Chennai, Tamilnadu, India
[*2, 3, 4, 5] UG Scholars, Panimalar Engineering College, Chennai, Tamilnadu, India

Abstract

Mental health is one of the most important things for a human life. In today's rapidly evolving electronic environment, maintaining mental health is crucial for both security reasons and hospital operations. This project's primary goal is to help mentally ill individuals by employing convolution neural networks (CNN) and received signal intensity indicators(RSSI). Processing statistical data is the approach taken in this case to effectively monitor behaviour. In this project, a system for wireless communication, as well as a person detection and tracking system, are the two main processes. As a result of this project, we are monitoring the behaviour and tracking of the patient more accurately using CNN and RSSI.

Keywords: RSS; CNN; Detection; Aggressive behavior; Alert.

1. Introduction

RSSI, or Received Signal Strength Indicator, is a telecom term. The RSSI is a tool used to gauge the power of the received radio signal. In most cases, RSSI won't be visible to the receiver. Since, because strength of the signals may vary in wireless networking, devices always provide the availability of measurements to the users. RSSI are always deduced in the intermediate frequency (IF) stage of the amplifiers. It is derived prior to the base band amplifier in the base band signal chain for zero-IF systems. After any potential cable loss, RSSI is a measure of the power level being received. The signal will therefore be stronger the higher the RSSI number. So, the closer the RSSI number is to 0, the stronger the received signal will be when it is delivered in a negative form, such as -50.

When the amount of radio energy in the channel is below a predetermined threshold value and the network card is clear to transmit (CTS), RSSI may be utilized internally to determine when a packet of data should be sent. As an illustration, Cisco system cards are said to offer 101 different power levels and have a maximum RSSI value of 100. The range of the RSSI value is 0 to 100. The way the 802.11 RSSI metric has been sampled results in one of its subtleties. RSSI do not require additional hardware necessities and it is present in all wireless nodes, for the purpose of localization, RSSI serves as the most possible indication device. The working of received signal indicator is shown in figure 1.

A considerable amount of attention has been gained by the behaviour analysis study of the user. For an instance, the hot areas or the hot spots of a room are determined through back/foreground detection. Techniques of Deep Learning are made on the bases of CNN (Convolution Neural Network); they are un programmed but they are only trained, so they need very least training which is statistically formal. This particular study proposal is the aggressive behaviour system based of the algorithm of CNN. And, here we use the software of MATLAB to run the CNN algorithm. In this system, the images used as input are captured in successive frames. By correcting for perspective, we should first convert the image to a 2D image in top view. And after that, we must proceed through detection, which will in fact involve both the tile approach and a mask-region based on CNN.

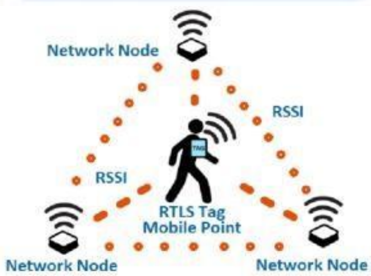

Figure 1: Working of Received Signal Strength Indicator

Finally, we would likely be storing all of the results in a cloud database, allowing us to create new heat maps based on 2D image.

Figure 2 displays full convolution neural networks.

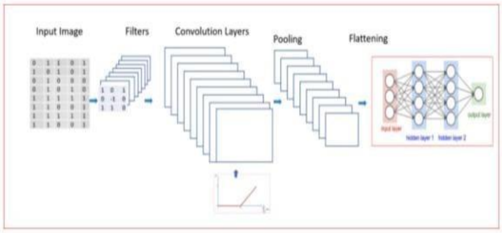

Figure 2: Full Convolution Neural Networks

2. Literature survey

The LAN-droids project aims to create a wireless mesh network made out of little, inexpensive robots. This robot network supports the tethering operations of the robot and serves as a platform for sensing and communication (I.e., it is the task of robot to localize, tracking and moving target to provide its continuous network coverage). There are a number of reasons to fix the LAN-droids tethering issues, including: initialization at random and absence of domain map.

The LAN-droids robots' responsibility is to offer a prior map, which is positioned in random places with random orientations. The robots were not aware of each other's relative locations and orientations because they shared a common reference point.

LAN-droids Robots has a limited sensing ability because of their low-cost design and it provides only a 2.4GHz wireless radio, odometer feedback, a push button sensor and just one short range IR. A data driven probabilistic approach to need a received signal strength(RSS) in robots for localization is introduced to solve the problem.

The requirement of LAN-droids project to automatically localize, track and follow the ill-equipped robots or humans in strange environments. Localization and tethering approach need the wireless signal strength and odometer without the crucial benchmarks for the field. For the tethering job, a grid-based localization technique is utilized, together with a data-driven probabilistic model for the distribution of distance and real-world received signal strength (RSSI). Only 2 nodes are proposed in this system. Multiple trackers can be added (Stefan et al, 2010).

Sports, health, and many other industries rely heavily on wireless sensor networks (WSNs), which are also employed for accurate motion and position tracking. In sports, tracking systems provide information on how the body moves during competition, while in medicine they are used to analyse motion anomalies and to treat wounded patients in clinics. The tracking system's precision has a big impact on performance.

It is suggested to use Received Signal Strength indicator (RSSI) readings from WSNs as the foundation for a new tracking system. The suggested method is made for WSNs that are installed near together and have a high data transmission rate. Spatial and temporal correlation between successive RSSI measurements is made possible by a sufficient transmission rate. Modern statistical and signal processing techniques are used to reduce channel distortion and make up for packet loss. Sensor nodes are not accurate, when used in real time. The calibration process must be changed because the accuracy delivered by a device under test is not same with that of standard of known accuracy (Gaddi et al 2013).

We develop a relative without an anchor localization algorithm for multi robot teams. The localization performs the collected readings based on the nodes' messages exchanged using Received Signal Strength Indicator. In order to give velocity vectors for each node in the network, RSSI measurement developed a relative velocity estimation framework and utilized previous distance measurement and location estimates. To obtain smooth RSSI pair wise signal distance for all nodes, they advise using the Floyd-Warshell algorithm and a Kalman filter. For comparative positions to pairwise distances, multidimensional scaling is obtained. Due to the lack of an anchor, relative positions are modified via geometric transformation to reflect the on-going motion.

The effects of various parameters to calculate the time it takes for the matrix to spread out, the use of synchronization and filtering of the RSSI data transmissions. This causes High exchange of wireless communications. Longer time is needed in order to converge to the final co-ordinates (Luis et al 2014).

Markets and international development encourage people to develop consumer habits that cause theft and robbery by consuming other people's possessions but not their own. Technological advancements give birth to the monitoring devices like CCTV. This allows us to keep an eye on the situation in the room and stop crimes before they happen. The cost of the equipment prevents everyone from purchasing it, and these two monitoring backdrops are simple to construct. The home monitoring system modifies and adapts to the user's demands using wireless devices. The system's wireless technology will be more convenient and accessible from any location.

The advancement of closed circuit television (CCTV) technology makes it easier for us to watch the space, but owing to the pricey nature of the gadget, not everyone in society has access to it. It is important to create a new system utilizing the Open WRT operating system that has features common to CCTV systems and is reasonably priced. The wireless router's operating system will be installed and equipped with a number of features, including speakers, GSM modems, webcams, flash drives, etc. The outcome of the system ability to track motion and record photos and videos of any suspicious movements. Also, it offers a number of capabilities that may be quickly activated via Wi-Fi and the internet, including alarm alerts, SMS warnings, user email reports, and alarm alerts. Maintenance of the system is required for whole devices (Arif et al 2015).

Low maintenance costs, accurate human localization in interior applications, and use of resources for security and health care are only a few of the essential applications. The other room occupancy-related aspects, such as lighting, water use, and controlled temperature, can all be lowered with their assistance. According to estimate from the health care industry, there will be approximately 2.1 billion older people worldwide in 2050, more than double the amount in 2015. By 2050, the proportion of people in working age to those over 65 is predicted to fall to 3.5. Parkinson disease can be identified in its early stages by observing behavioral changes in human behaviour.

Indoor person localization does not have accurate tag and it has some applications like health monitoring and supported living. Sensor data variability and noise can be reduced using machine classifiers as a result of deployment-specific environmental variables. This study examines the effectiveness of Weka collection ML classifiers using experimental data from using capacitive sensors for indoor human localization in a 3m x 3m room. The variance resources needed by the algorithm for training and inferring, as well as the size of the training set are compared with the algorithm's localization. The output of ML classifiers demonstrates a significant variation in method, accuracy, precision, and recall surpassing 93% and a localization error of 0.05 m on average.

This article's primary contribution is an examination of the Weka set of ML classifiers performance in localizing capacitive sensors with the least amount of pre-processing. We must quickly go over capacitive sensor design and applications in order to keep an eye on the constrained space. The experiment's findings demonstrate a significant improvement in localization precision, memory, and accuracy compared to our earlier work. How to evaluate the size of the training sets for the best-performing and improved classifiers is demonstrated. Room is 3m*3m only for experiment purpose. The quality of the result is not perfect. (Osama et al 2017).

To analyse the behaviour of the students, two potential approaches are taken. They are surveys and quizzes. It is currently not difficult to design computer vision-based activities for the students to analyse their work and classroom behaviour. The objective of this study is to create an automated system that will faculty to monitor and summarize student's classroom behaviour. They are utilized by this data in the decision-making process. The computer keeps track of every class period and notes when the kid pays attention. After analyzing the all data, the system prepares the report to the faculties. The system's advantages include being more precise and adaptable than the earlier research.

This research designed to work the system automatically to supports the teachers and educational related faculties to monitor the student behaviour. With its automatic provision of strategic information, the system serves as a judgement call helper. This carries out the student behaviour records, carrying out statistics, and displays the data. In this experiment, we demonstrated the ability to combine data and resolve the student behaviour tracking conundrum. This keeps track of the absence of other pertinent data, such feelings. Architecture processing systems are (Bui et al 2019).

Using deep learning, it is possible to identify and visualize the hot spots that are present while also estimating the amount of individual's currently present real time in retail establishments. It requires RGB camera (i.e., video recording device) is inexpensive. To fix and solve the problem of counting people, employ supervised learning techniques based on a CNN regression model. RGBP image has four channel of binary picture and a standard RGB image. P stands for whether a person can be seen in any given pixel of the image.

A foreground/background identification system that takes into account the peculiarities of people's activity in retail outlets is created to decrease the final information. To identify the store's hot zones, which can be studied, the P image is used. Several experiments were carried out to validate, assess, and contrast our strategy using data set contains videos which is collected from surveillance camera in retail store. The result for our approach is adequately well employed in real-world scenarios and does simple CNN procedures.

A suggested Convolution Neural Network (CNN) regression model is employed to resolve the counting issue via supervised learning, each region of the spatial domain's population density can be discovery by; CNN uses four channel pictures with the abbreviation RGBP (red, green, blue, people). This notation drift-based real-time foreground detection approach is proposed. RGB videos can be used for a foreground detection method.

The colour and foreground information provided by an input image format (RGBP). In order to estimate the population of an image using standard RGB or P photos, a CNN regression model trained and practiced counting individuals from the aforementioned RGBP images. This improved its estimation accuracy. A training model cannot be used elsewhere or view point the similar locations but various training is tailored to each location and perspective. Extrapolation do not support trained models for high non linearity applications. (Valério et al 2019).

3. Existing System

The widespread of concern people had been focus on the public safety issues from lifework. The development of video detection technology is essential for preventing problems with public safety, which can identify anomalous human activity in videos. The framework to examine the behaviour of unstable person's indoors is detected using this abnormal behaviour of indoor peoples and a new abnormal behaviour. Figure 3 displays the block diagram of the current system.

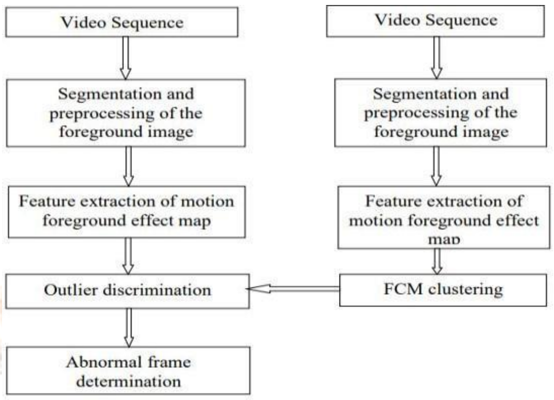

Figure 3: Current system's block diagram

The block diagram illustrates the detection of aberrant behaviour in indoor crew members and makes use of a new framework for detecting abnormal behaviour. First, a background modelling technique based on a Gaussian mixture model is utilized to separate the backgrounds of each image frame in the film. Following background segmentation, block processing is utilized to produce the space-time block of each frame image, which is also employed as the basic illustration of the detection object. The foreground picture components of each space-time block are extracted in the third block. The fourth block makes use of fuzzy C means clustering (FCM) to find outliers in the sample data. Video is converted into number of frames and these are used for the upcoming pre-processing stage. Pre-processing procedures are used at this step to pre-screen the raw video data. Foreground extraction and video data normalization are two techniques used to remove the unnecessary information from images and improve the ability to discern between important and unimportant details. Local temporal and spatial features and trajectory segmentation features are the features that make up the characteristics of crowd movement. In literature research, the majority of the local spatial temporal data are used to obtain the motion data and additional features that is recovered primarily consists of gradient and optical flow features. The foreground picture is integrated with the spatial block and pre-processed to extract the foreground's motion, which is then represented by motion in the intricate visual pattern video frame. A specific space time block's feature description in detail is obtained by averaging the successive multiple frames of space blocks after spatial block's for each block of foreground motion, effect weights have been generated are as follows:

A. Foreground Motion Block

The building where the space is picturesque place first emerges is known as the foreground motion blocks. The foreground motion block comprises information in some spatial blocks, such as noise, which cannot accurately describe an object's motion behaviour. This study pre-processes all of the spatial blocks and establishes that such motion barriers exist.

B. FCM

Fuzzy C means Clustering algorithm which is based on distance. Here, it is given a dataset = {1, 2 … } which consisting of n p-dimensional samples and the data are grouped into (∈ [2,])categories. The centre of each category is s · = {1, 2 … } is the cluster center matrix. = $\{u\}$ ∈, which fits the criteria of membership $\sum un$=1 = 1. The following is how FCM actually works:

$$J = \sum_{k=1}^{n} \sum_{i=1}^{c} U_{ki}^{m} \|x_i - v_k\|^2 \quad (1)$$

C. Drawbacks

Machine learning algorithms is used for a detection framework. The proposed detection framework needs many parameters manually and this is the drawbacks of the research.

D. Problem Identification

Detecting the behaviour of human being using video sequencing has an important factor accuracy. Accuracy has reached only 90% which increased in the current system.

4. Proposed System

Within the proposed system, human movements and their location inside a closed place or public place are being tracked in terms of zones. Number of electronic nodes are set up in different places inside a building so that it tracks human movements who have personal assistant like mobile phones. In this paper, a device-free person detection and tracking system for a specific parameter is proposed using Received Signal Strength Indicator.

A person tracking and person-detection system, as well as wireless communication are the two main components of the suggested system. The first is designed to measure and gather RSSI signals brought on by the mobility and presence of people, whereas latter uses a predetermined threshold and a zone selection algorithm to recognise and track the person.

The suggested system is novel in that the communication protocol can prevent network signal interference and packet loss, and the detection tracking method can identify a precise area where the human is present by taking into account an ideal predefined threshold and a level of RSSI variation. There are three zones which are divided into green zone, orange zone and red zone. The green zone denotes good (safe) zone. The orange zone denotes fair. The red zone denotes poor (out of the building). Experiments and studies of human movement patterns at various speeds and orientations are used to test and validate the suggested system. The experimental finding suggests that suggested communicating procedures can offer exceptional communication dependability, assuming that offered approach is viable as a result detect and track human motions. The reliability of the communication is shown by delivery rate for packets, which is nearly 100%. When one guy moves, accuracy of detection and tracking, as measured by the percentage of times the method can do so with right zone, is around 100% in all circumstances.

We also propose the use of CNN technique for the detection of aggressive conduct in addition to the RSSI tracking system. The input image for this system is a series of video frames recorded by a monitoring camera. Perspective adjustment and person behaviour detection are the two key operations carried out by this technology. In order to create a perspective transform matrix, an input image is converted into a 2D top-view image using perspective correction. Users are able to operate the system more effectively thanks to it. The person behaviour detection algorithm generates a duplicate of the input image. On each tile image and the full-frame image,the CNN is utilised to detect objects. The bounding boxes for each tile image and the full-size images are kept as the initial results once the unstable person's conduct is initially detected using CNN. In order to provide a final detection result from the database, the first results are eventually combined utilising our suggested scheme. Additionally, we compute the behaviour coordinates of the person using the bounding boxes in the end product; using the perspective transform matrix, these are then transformed into new behavioural coordinates for the person in 2D photos. The mechanism can thus be immediately inspected from various angles. It may be difficult to count the number of individuals in the photo due to the camera angle, especially if they are little or in a crowd. To solve this issue, the perceptual correction approach is used to convert the input into a 2D top-view image.

A. Transmitter section of RSSI

Figure 4: Block diagram of Transmitter

In the transmitter section, the RSSI is connected to the mobile phone through the URL link which is shown when the hot spot of the mobile phone is connected. The location will be sent along with the alert message to the person in-change through GSM. Figure 4 depicts the transmitter's block diagram.

B. Receiver Section of RSSI

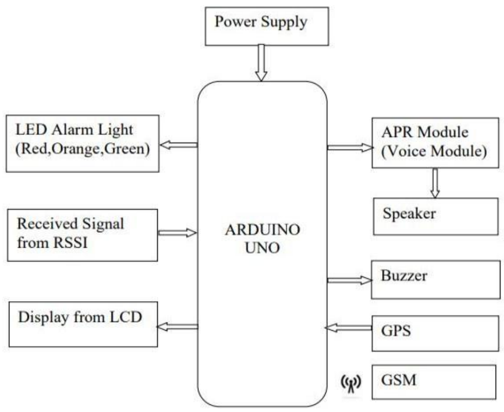

Figure 5: Block diagram of Receiver

In this system, the Arduino UNO micro-controller constitutes as brain of our system, in which the entire system program is stored in it. There are three zones in the parameter: green(safe) zone, yellow(fair) zone and red(poor) zone. If the patient escapes from the building (red zone), it will alert the nurse/warden. As soon as the patient cross the red zone, the LED LIGHT will be turned on and buzzer is used to alert. Furthermore, the voice module speaker will be turned on for alerting and LCD display will indicate the zone. Even after they cross the red zone, the patient can be tracked through GPS. Figure 5 displays the transmitter's block diagram.

C. Behaviour Detection

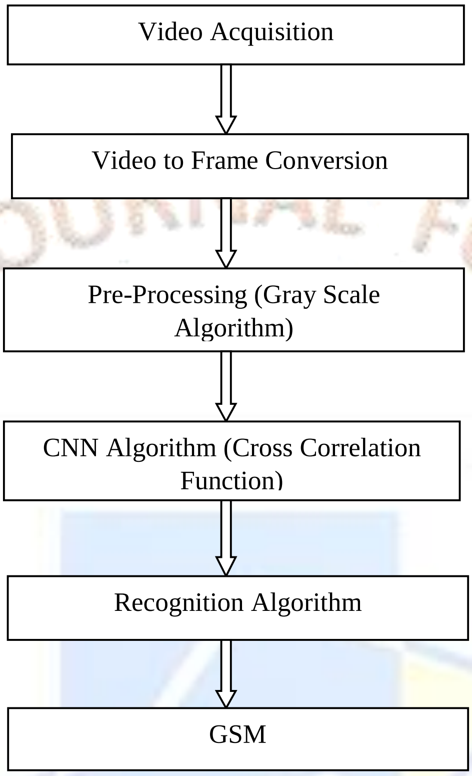

Figure 6: Flow chart of Behaviour detection

The input image is processed from the video of mentally unstable person's behaviour in hospitals. The video is transformed into sequence of images/frames. The generated input image is sent into the pre-processing section where the noise is eliminated and re-sized. Then the image enters into the CNN layers. The image is filtered by the layers of CNN and the recognition is done by CNN algorithm. The output image is classified and the detected behaviour will be sent to the nurse/warden by GSM. Flow char of behaviour detection is shown in figure 6.

D. CNN ALGORITHM

The Convolutional Neural Networks (CNN) are among the most well-known deep learning techniques and are frequently employed in image categorization applications. In the CNN architecture, the three primary layer types are convolutional, pooling and fully connected layers. The CNN algorithm takes an input image as input, processes it via the layers to recognize and identify features, and then outputs a classification. Convolutional layers and pooling layers are alternated in the CNN's design, followed by a group of fully linked layers. Each layer in a CNN's output becomes its subsequent layer's input.

a. Image input layer

Specifying the image size, in the example 28by 28 by1, is done through an image input layer. The channel size, height, and width are represented by these numbers. The colour channel's size is 1, as the digit data is made up of gray scale images. The RGB values are used to determine the channel size for a colour image, which is 3. You don't have to because the train network shuffles the data automatically when training starts, Additionally, Train Network has the ability to automatically shuffle the data at the beginning of each training epoch.

b. Convolution Layer

The first argument in the convolutional is filter size, which specifies the height and width of the filters that are applied by the training function as it moves over the images. The number 3 in this instance denotes a 3-by-3 filter size. The second argument, number Filters, is the number of neurons that connect to the same region of the input as filters. The quantity of feature maps is set by this parameter. The input feature map can be given padding by using the 'Padding' name value pair. With 'same' padding, the spatial output size is guaranteed to be
the same as the input size for a convolutional layer with a default stride 1. In addition, you can specify this layer's stride and learning rates sing the name-value pair options of the convolutional 2d layer.

c. Batch Normalization Layer

Batch normalization layers normalize the activations and generates spreading through a network to make network to make network training an easier optimization problem Utilize batch normalization layers between convolutional layers and nonlinearities, such as ReLU layers, to speed up network training and reduce the vulnerability to network initialization. For the purpose of creating a batch normalization layer, use batch normalization layer.

d. ReLU Layer

An activation function that is nonlinear comes after the batch normalization layer. Rectified linear unit (ReLU) activation functions are most frequently used. Make a ReLU layer by using ReLU Layer.

e. Max Pooling Layer

Down sampling, which is sometimes applied after convolutional layers, reduces the spatial size of the feature map and eliminates unnecessary spatial information. Deeper convolutional layers may include more filters thanks to down sampling because the amount of processing needed for each layer remains constant. Utilising a produced technique for down-sampling is max pooling, which is provided by the maxpooling2dlayer. The first-option, pool size, specifies a rectangular input region whose maximum values are returned by the max pooling layer. ‘stride’ names and values option describes the size of the training’s step functions uses while it reads the input.

f. Fully Connected Layer

The convolutional and down-sampling layers are followed by one or more fully linked layers. As the name implies, a fully linked layer contains all of the neurons in the layer above it. This layer incorporates all the knowledge that the other layers have acquired in order to identify the broader patterns in the image. The final, completely integrated layer combines the attributes to classify the images. As a result, the Output Size parameter in the final fully connected layer's output size parameter matches the number of classes in the target data. In this instance, the output size is 10, which matches the number of classes. Use fully Connected Layer to create a fully connected layer.

g. Softmax Layer

The softmax activation function adjusts the output of the fully linked layer. The output of the softmax layer is a positive value that sums to one, which the classification layer can use to calculate classification probabilities. Use the softmax Layer function to construct a softmax layer after the last entirely linked layer.

h. Classification Layer

The final layer is categorization. This layer uses the probabilities provided by the softmax activation function for each input to classify each input into one of the mutually exclusive classes and compute the loss. Working of CNN algorithm is shown in figure 7.

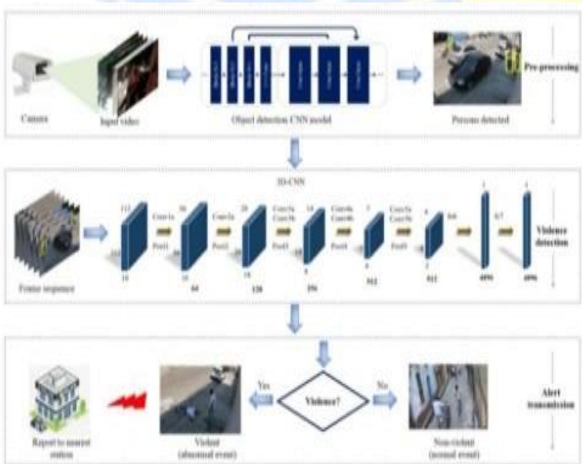

Figure 7**:** Working of CNN algorithm

i. Function used in CNN

Convolution is a mathematical procedure that uses input I and kernel K as arguments to create an output that explains how the shape of one is altered by the other.

A feature map is produced by applying a mathematical operation on a image (“x”) that is a 2D array of pixels with distinct color channels (Red, Green, and Blue-RGB) and a feature detector or kernel (“w”).

$s[t] = (x * w)[t] = \sum_{a=-\infty}^{a=\infty} x[a]w[a+t]$ (2)

Convolution is represented in equation 2. The computation of the mathematical procedure determines how similar two signals are. Convolution operations will assist in detecting the edges in the picture when we apply a feature detector or edge-identifying filter to an image that may contain an edge-identifying feature detector or filter. It is possible to implement the infinite summation as a sum over a finite number of array elements.

$$s(i,j) = (I * K)(i,j) = \sum\sum I(m, n)K(i - m, j - n) \quad (3)$$

Where, 2D array I and kernel-Convolution function K.

The rewritten version of equation 2 is equation 3 because convolution is commutative. Due to the narrower range of possible values m and n, this makes implementation simpler.

$$s(i,j) = (K * I)(i,j) = \sum\sum I(i - m, j - n)K(m, n) \quad (4)$$

- The equation 4 represents the cross-correlation function.
- Convolutional layer is a fundamental component of CNN and aids in feature detection.
- Kernal K is a collection of trainable filters that spans the entire depth of the input image despite being spatially small in comparison to the image. We are aiming to detect features and construct several feature maps to assist us identify or categorize the image. Kernal K, a feature detector, is analogous to the torch on image I.

To assist with tasks like edge detection, recognizing various forms, bends, or colors, etc., there are numerous feature detection available.

5. Results and Discussions

The accuracy of the aggressive behaviour is higher than machine learning techniques in CNN. The result of accuracy of aggressive behaviour of patient with given dataset is shown in the figure 8 and the result of accuracy of normal behaviour of patient with the given dataset is shown in the figure 9.

Figure 8: Output of Aggressive behaviour

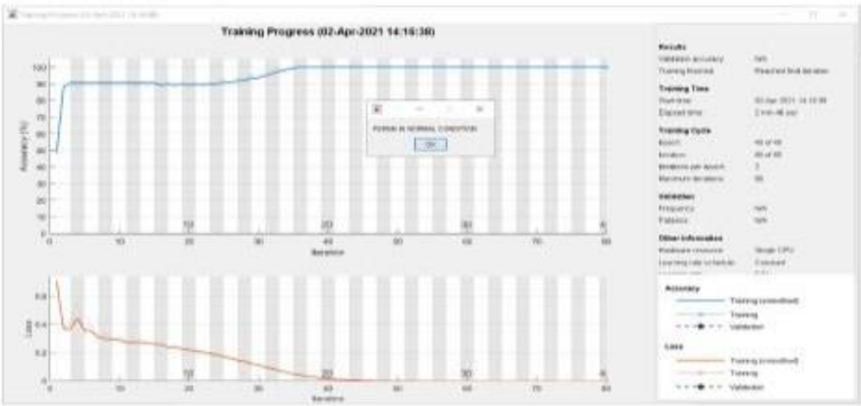

Figure 9: Graph of Normal Behaviour

Along with graph, the alert will be given to the nurse as shown in figure 10 and this is provided with the help of hardware part of GSM module as shown in figure 11 helps to save the patients from the patient who is aggressive and helps to treat the patient accordingly.

Figure 10: Message of Patient Behaviour

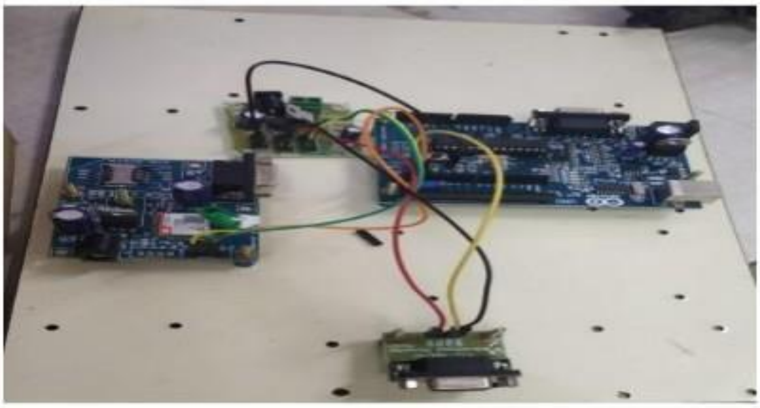

Figure 11: Hardware implementation of image processing.

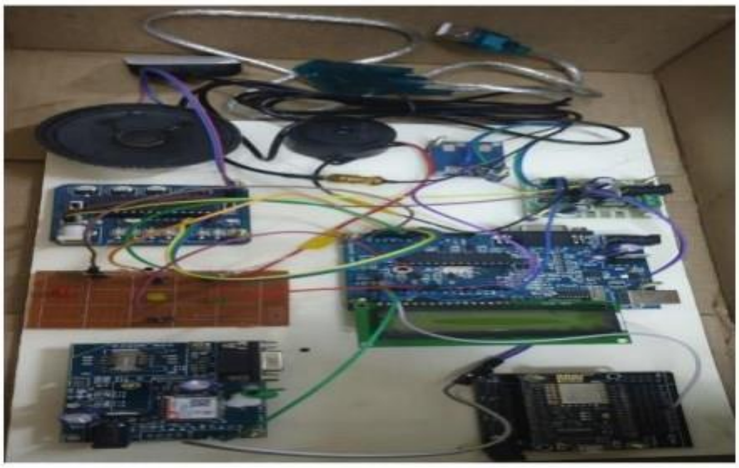

Figure 12: Hardware implementation of RSSI

In the hardware implementation of RSSI, the patient safety zone will be monitored and the alert message will be sent according to zone as shown in figure 12.

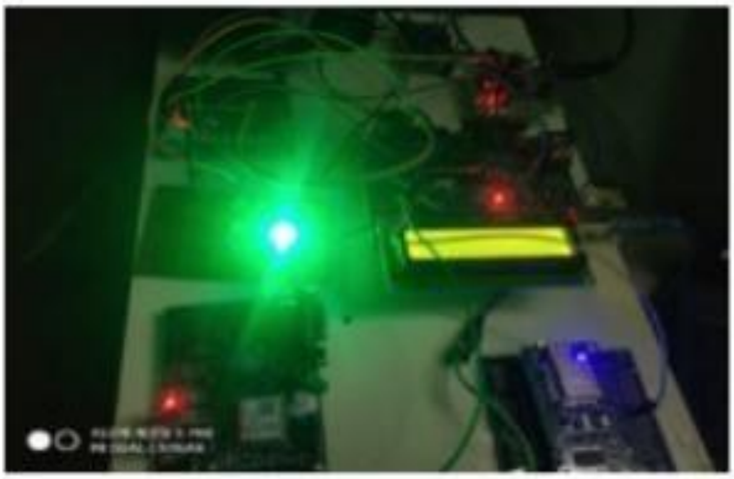

Figure 13: Safe zone

Figure 14: Mid zone

Figure 15: Red zone

From the above images, it is known that figure 13 represents the patients are in safest zone with the help of green light. Figure 14 represents that the patients are in mid zone with the help of yellow light and the figure 15 shows that the patients are in red zone with the help of red light.

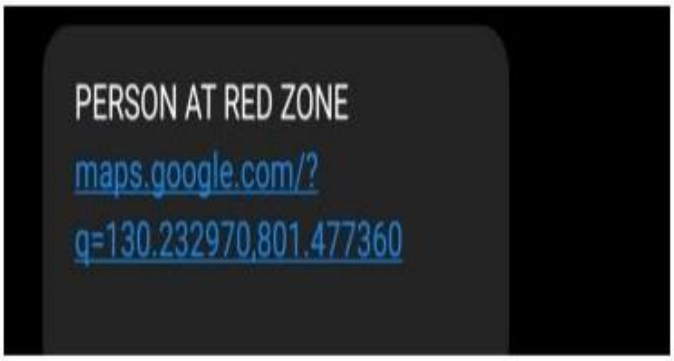

Figure 16: Message for alert

The Figure 16 represents the message received if the patient has escaped from the red zone. With the help of GPS, the patient can be tracked if escaped.

6. Conclusion

In this paper, the proposed method is a pure radio frequency according to the relative positioning and reaches the system to reach the radio frequency emitting object autonomously. The app uses the user's estimated distance, angle and speed as the antenna rotates. We demonstrate our system's performance and error statistics through simulated process and real world experiments using our models. We developed a method for detecting hostile behaviour utilizing deep learning and perspective-correction techniques in combination to the surveillance system. There are many research questions that need to be addressed in our future work. By using the concept of compressing the instance the system can be made faster, which may allow shortening the time as I continue. Future security checks may be requested from to learn more about our recommended checks. By implementing a parallel system, this system's uptime can be increased while data is still being gathered and analyzed for the system. Project intended to visit additional locations, including shops and supermarkets.

References

[1] P.Barsocchi, S.Lenzi, S.Chessa, and G.Giunta, A novel approach to indoor rssi localization by automatic calibration of the wireless propagation model, in IEEE VTC 2020.

[2] J. Graefenstein, A. Albert, P. Biber, and A. Schilling, Wireless node localization based on rssi using a rotating antenna on a mobile robot, in IEEE WPNC, 2021.

[3] S. Zickler and M. Veloso, Rss-based relative localization and tethering for moving robots in unknown environments, in IEEE ICRA, 2020.

[4] J.R. Jiang, C.M. Lin, F.Y. Lin, and S.T. Huang, Alrd: Aoa localization with rssi differences of directional antennas for wireless sensor network, in International Conference on i-Society, 2019.

[5] G.Blumrosen, B.Hod, T.Anker, D.Dolev, B.Robinsky, Enhanced calibration technique for rssi-based ranging in body area networks, Ad hoc networks, vol. 11, no. 1, pp. 555–569, 2018.

[6] J. Xiong and K. Jamieson, Array track: a fine-grained indoor location system, in USENIX NSDI, 2021.

[7] G. Blumrosen, B. Hod, T. Anker, D. Dolev, and B. Rubinsky, Enhancing rssi-based tracking accuracy in wireless sensor networks, ACM Transactions on Sensor Networks (TOSN), vol. 9, no. 3, p. 29,2023.

[8] L. Oliveira, H. Li, L. Almeida, and T. E. Abrudan, Rssi-based relative localization for mobile robots, Ad Hoc Networks, vol. 13, pp. 321–335, 2022.

[9] Setiawan, A., Yazid, A.S.,Wahyudi, M.D.R. Room monitoring system using open WRT-based webcam, Int. J. Inform. Dev.. 4, 15–23, 2018.

[10] He, K., Zhang, X., Ren, S., Sun, J., Deep residual learning for image recognition, in Proceedings of the 2016IEEE Conference on Computer 61 Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30, pp. 770–778, 2021.

[11] Tariq, O.B.; Lazarescu, M.T.; Iqbal, J.; Lavagno, L., Performance of machine learning classifiers for indoor person localization with capacitive sensors, IEEE Access, pp. 12913–12926, 2017.

[12] Karnalim, O.; Budi, S.; Santoso, S.; Handoyo, E.D.; Toba, H.; Nguyen, H.; Malhotra, V., Face - Face at classroom environment: Dataset and exploration, in Proceedings of the 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), Xi'an, China, pp. 1–6, 2021.

[13] Nogueira, V.; Oliveira, H.; Silva, J.A.; Vieira, T.; Oliveira, K., Retail Net: A deep learning approach for people counting and hot spots detection in retail stores, in Proceedings of the 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Rio de Janeiro, Brazil, 28–31, pp. 155–162, 2019.

[14] Anh, B.N.; Son, N.T.; Lam, P.T.; Chi, L.P.; Tuan, N.H.; Dat, N.C.; Trung, N.H.; Aftab, M.U.; Dinh, T.V., A computer-vision based application for student behavior monitoring in classroom, Appl. Sci., 9, 4729, 2019.

[15] Dow, C.R.; Ngo, H.H.; Lee, L.H.; Lai, P.Y.; Wang, K.C.; Bui, V.T., A crosswalk pedestrian recognition system by using deep learning and zebra crossing recognition techniques, Softw. Pract. Exp., pp. 1–15, 2020.

[16] Jiang, B.; Xu,W.; Guo, C.; Liu,W.; Cheng,W., A classroom concentration model based on computer vision, in Proceedings of the ACM Turing Celebration Conference- China (ACM TURC 2019), Chengdu, China,17– 19, pp. 1–6, 2019.