

Personalized Health Framework for Visually Impaired

Megha Rathi, Shruti Sahu, Ankit Goel and Pramit Gupta

E-mail: megha.rathi@jiit.ac.in, shrutisahu1196@gmail.com, goelankit1995@gmail.com, pramitgupta22@gmail.com
Department of Computer Science & IT, Jaypee Institute of Information Technology, Noida, India

Keywords: android application, computer vision, deep learning, object recognition, region-based convolutional neural networks, disease prediction, voice assistant

Received: August 23, 2019

Vision is one of the most essential human sense. The life of a visually impaired person can be transformed from a dependent individual to a productive and functional member of the society with the help of modern assistive technologies that use the concepts of deep learning and computer vision, the science that aims to mimic and automate human vision to provide a similar, if not better, capability to a computer. However, the different solutions and technologies available today have limited outreach and end users cannot fully realize their benefits. This research work discusses an easily-operable and affordable android application designed to aid the visually impaired in healthcare management. It also aims to resolve the challenges faced due to visual impairment in daily life and uses the concepts of computer vision and deep learning. Broadly, the application consists of the following modules: object recognition in immediate surroundings using region-based convolutional neural networks, disease prediction with the help of symptoms, monitoring of health issues and voice assistant for in-app interaction and navigation.

Povzetek: Razvita je androidna aplikacija za pomoč ljudem z okvarami vida.

1 Introduction

In 2014, the World Health Organization estimated 285 million people to be visually impaired worldwide, out of which 39 million are blind and 246 have low vision [1]. About 90% of this population lives in low-income settings. Visually impaired people face several difficulties in their daily lives. From reading to navigation, be it in familiar or unfamiliar environments, every task is a new challenge [2]. Computer Vision is the science that aims to mimic and automate human vision and provide a similar, if not better, capability to a machine or computer [3]. With the combination of machine learning and computer vision, various technologies have been developed to help substitute for visual impairment in some manner or the other and enable people to live more independently. However, the problem of lack of accessible and affordable solutions to ease the routine of the visually impaired still persists. Expensive wearable's, usually powered by artificial intelligence, are the current advancements in computer vision. Aiming towards affordability, a simpler approach is adapted. Android platform is used to develop an application and make it accessible in hand-held mobile android devices which are essential in developing an aid for the visually impaired [4].

The objective of the proposed application significantly in facilitating an independent & healthy life include easy operability i.e. voice assistant for in health routine, object recognition in images captured via device camera, diagnosis of common diseases via symptoms spoken by the user. The modern digital era has revolutionized data storage. Huge volumes of data are

available today that can be beneficially utilized for processing and automation an essential feature of health management is incorporated in the android application, along with health monitoring measures like body mass index (BMI), calorie intake, and daily steps. Before the breakthrough of deep learning in 2000s, PASCAL and conventional computer vision techniques like example-based learning, discriminatively trained part were used for object recognition structure as early deep learning based algorithms (for instance, R relevant regions, labeling each proposed region and cumulating outcomes from all output for the image [5] Figure 1 illustrates the classification error in top 5 models of the object recognition & image classification task of ImageNet Large Scale Visual Recognition [6]. It can be observed that deep learning even surpassing human error in 2015. A deep learning Region-based Convolutional Networks (R 'Methodology') is used to detect and identify objects from the images captured by the android device [7]. Figure 1. Classification Error of top 5 models of ILSVRC application is to provide an affordable solution significantly in facilitating an independent & healthy lifestyle for the visually-impaired voice assistant for in-app interaction and navigation, management of daily health routine, object recognition in images captured via device camera, diagnosis of common diseases The modern digital era has revolutionized data storage. Huge volumes of data are available today that can be beneficially utilized for processing and automation [8]. Automation of disease prediction is an essential feature of health management is incorporated in the android application, along with health mass index (BMI), body fat percentage (BFP) and basal metabolic rate intake, and daily steps. Before the breakthrough of deep learning in

2000s, PASCAL and conventional computer vision based learning, discriminatively trained part-based models and selective search [9]. Various algorithms shared similar structure as early deep learning based algorithms (for instance, R-CNN) i.e. identification and proposal of each proposed region and cumulating outcomes from all regions to produce. Illustrates the classification error in top 5 models of the object recognition & image classification Large Scale Visual Recognition Competition (ILSVRC) between 2010 and 2016. It can be observed that deep learning-based models fare much better than others, even surpassing human error in 2015. Salient features navigation, management of daily health routine, object recognition in images captured via device camera, diagnosis of common diseases the modern digital era has revolutionized data storage. Huge volumes of data are available today that can. Automation of disease prediction is an essential feature of health management is incorporated in the android application, along with health and basal metabolic rate before the breakthrough of deep learning in 2000s, PASCAL and conventional computer vision based models and selective search. Various algorithms shared similar CNN) i.e. identification and proposal of regions to produce illustrates the classification error in top 5 models of the object recognition & image classification even 2010 and 2016 based models fare much better than others, (elaborated under ‘Methodology’) is used to detect and identify objects from the images captured by the android device.

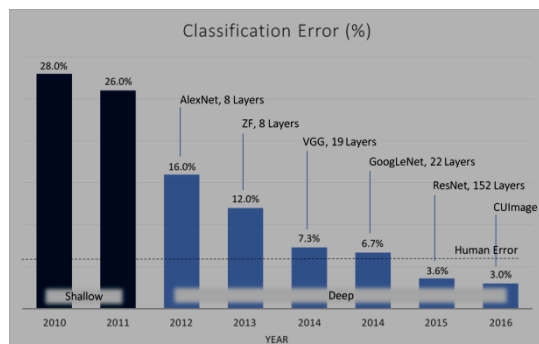


Figure 1: Classification error of top 5 models of ILSVRC.

2 Background study

Significant research has been carried out in the domain of developing android based apps for visually impaired patients. In recent research, authors have presented a novel technique for visually impaired user assistance using guidance mode activity [10]. A proposed system is comprised of gathering sensors data held or worn by a visually impaired user, these sensors collect data and further redirect it to the server for processing. The server contains a processor which is built using advanced artificial intelligence techniques. Basis functionality of the server is to send data extracted from sensors to an agent device, this device further gives data for vision on

an agent interface. Finally, an agent can assist the low vision user in real-time navigation through audio signals or other feedback.

In another research, a system named “Smartvision” is developed for the navigational activity of blind users. With the recent development in advanced computational technologies, one can create models which can assist blind users in their daily routine tasks. The main emphasis of the proposed model is to assist no vision users so that they can easily navigate into strange environments indoor or outdoor [11]. A user-friendly interface is developed for the development of “Smartvision” app for blind users. The proposed model utilizes the concepts of computer vision and machine learning to achieve the objective.

In yet another novel research work, an android healthcare app known as “mHealth” is developed to assist blind users in their health tasks. Android-based technologies are gaining popularity in healthcare these days and could serve as a boon for visually impaired patients [12]. Smartphones allow extracting medical data from health sensors and then using the extracted data for further health analysis of the patient. Visually impaired people have bad health conditions than normal people because of poor accessibility of medical data and if a compatible app is present they can regularly monitor their health status and take preventive action accordingly. “mHealth” device is proven to be a boon as it captures patient health condition via sensors i.e. it can monitor blood pressure, diabetes level, etc. and further suggest medical action accordingly. A smartphone is the simplest way to access health sensors, and android developers can create IoT-based android applications to make medical sensors fully accessible on mobile devices.

A study conducted in another significant paper developed a novel technique for door detection for the strange environment to no vision user [13]. Earlier algorithms were designed to detect door-like objects to known environments only where particular characteristics of the door are fed as an input. In this study, the author presented a novel image-based door detection technique whose main objective is to find outdoors based on consistent characteristics like edges and corners. Animated door structure is created for finding out the doors by merging features of edges and corners. This algorithm is also able to distinguish other door-like objects from the door like it can distinguish bookshelf or cabinet from the door. The proposed technique is validated under different unknown situations over multiple ranges of door shapes, colors, textures, illumination, and views.

Another contribution presented recent advances in advanced homecare technologies for blind people. Assistive systems are required for blind users to provide information and allow them to safely move and complete their daily routine and health tasks and explore unknown environment [14]. For achieving this objective various new IOT based and other computational technologies have been experimented with to provide the solution to basic major problems of a blind user. For blind people, accessibility and self-determination are elementary

requirements so it is desirable to provide them insight regarding skills to lead a happy normal life like other human creatures. For blind self-determination and accessibility means they can seek employment, get a good education, get normal health routines, and social life. So, this study focuses on providing details about various smart home care solutions for blind users.

An Integrated mobile-based healthcare framework is used these days by medical professionals for healthcare tasks [15]. The usage of smartphones is expanding day by day and in the future, it incorporates every single clinical task. An easy-to-use interface makes healthcare apps be used by even illiterate persons. Effective utilization of advanced computational techniques, proper verification, and validation are substantially required to ensure a good standard of quality, security, and privacy for using these mobile-based healthcare applications. With the execution of all such quality standards in mobile-based healthcare tools, the main emphasis is on providing correct, relevant, and appropriate information to the user for achieving the objective of healthcare outcomes.

Within this paper extra emphasis has been put on enhancing the assistive technologies used for visually impaired people [16]. Assistive technology is used by many researchers to assist blind users but academicians are not paying attention to deriving new applications by amalgamating assistive technologies with computational intelligence for creating hybrid applications to assist the blind user. The socio-psychological feature is the main obstacle in the adoption of assistive technology for further research. Finding and pointing out those features can enhance the adoption of assistive technology by academicians. Visually impaired patients are heavily relying on these technologies for survival, this research focuses on finding out the socio-psychological feature that impacts assistive technology.

Another significant contribution provided in the study, in which authors generate particle libraries for drug discovery using recurrent neural networks [17]. The recurrent neural network can be proven to be very effective for creating systems for molecular structures, features of molecular structures coordinate positively with the features used to train the model. The proposed work proposed a model with a tiny molecular dataset that is dynamic in opposition to the target. The proposed model can outline the method for producing a huge set of molecules for discovering drugs.

Research conducted the study over a survey of Wearable Obstacle Avoidance Electronic Travel Aids for Blind [18]. Numerous wearable devices are developed to assist blind people in navigation in a known and strange environment. Broadly these navigational devices are classified into three categories: automatic inclination aid, electronic progression aid, and posture location tools. This research work presents the survey of portable navigational devices for visually impaired patients. This can help researchers in gaining insights to present assistive technologies for blind people and further amendments in these technologies.

Recent developments in computer vision include expensive wearables like Aira, MyEye, eSight, and BrainPort. Inspired by Google Glasses, Aira, developed is a pair of spectacles fitted with a camera that transfers the current field of view to a visually-abled person i.e. provides visual interpreting services [19]. Unlike Aira, artificial intelligence is used in MyEye, launched by which interprets visual data from a small camera into an audio earpiece. eSight, designed by Conrad [20] for the partially blind, uses a high-resolution camera to enlarge images and project them on an OLED screen in front of the wearer's eyes.

Object recognition and image processing are major functionality of computer vision systems. Deep neural networks, an arrangement following deep learning are networks with multiple hidden layers. However, as the depth of the network grows and it begins to converge, accuracy becomes constant and then reduces rapidly. A residual learning framework, ResNet, to easily train substantially deeper neural networks was presented by He et.al. [21]. Consider two neural network layers x and y . Instead of direct mapping $y=F(x)$, ResNet adapts a residual function $F(x)$ to map $y=F(x)+x$. The logic is that in the case of identical layers $y=x$, it is simpler to obtain $F(x)=0$ than $F(x)=x$. Hence, layers are framed by learning residual functions.

The author proposed a combination of convolutional neural networks with region proposals, called Region-based Convolutional Networks (R-CNN) [22]. The model inputs an image proposes bounding boxes or region proposals using selective search and checks if each proposal is an object or not. A Support Vector Machine (SVM) is used for the classification of region proposals or bounding boxes which are tightened using linear regression. The architecture has significantly better results than earlier CNN-based architectures on datasets like ImageNet.

Proposed by Girshick et.al. [23], Fast R-CNN is a simplified version of R-CNN. It clubbed all the components of R-CNN into a single network by adding a softmax layer and a linear regression layer parallel to the output layer of CNN. Softmax acts as a classifier in place of SVM and linear regression tightens the bounding boxes. In the same year, Faster R-CNN was proposed by Ren et.al. [24] to accelerate the region proposal process. The model trained a single CNN to implement region proposals and classification by adding a fully convolutional network, named Regional Proposal Network (RPN) on top of CNN. Anchor boxes are some common aspect ratios that are randomly generated to fit objects in an image. RPN slides a window over the CNN feature map outputs a bounding box per anchor and probability if it contains an object. These boxes are then passed to Fast R-CNN for classification.

A feature pyramid is a fundamental component in recognition systems for detecting objects at different scales which is troublesome for tiny objects. The proposed work [25] used the framework of convolutional neural networks to construct feature pyramids. This architecture, called a Feature Pyramid Network (FPN), comprising of a bottom-up (downscale) and a top-down

Year	R-CNN	FastR-CNN	Faster R-CNN
Year of Conception	2014	2015	2016
Input	Image	Image with region proposals	Image (region proposals not needed)
Output	Bounding boxes and labels for each object in the image.	Object classification of each region with more constricted bounding boxes.	Classifications and bounding box coordinates of objects in the images
Components	CNN (feature extractor), SVM (classifier), linear regressor (tighten bounding boxes).	Single CNN (feature extractor) having softmax layer (classifier) with linear regression layer (output bounding boxes) in parallel.	CNN (feature extractor), RPN (output bounding boxes per anchor and probability), Fast R-CNN (classifier, tighten bounding boxes)
Pooling	Max Pooling	RoI (Region of Interest) Pooling	--
Region Proposal	Selective Search	Selective Search	Region Proposal Network

Table 1: Comparison of existing approaches for object recognition.

(upscale) pathway. The bottom-up pathway uses ResNet for construction. Upsampling in the top-down pathway is done by convolutional filters. FPN shows significant improvement as a feature extractor in several applications.

In another recent study, for object detection deep convolutional neural networks (CNNs) are trained as N-way classifiers [26]. HMod Fast R-CNN is implemented for upgrading the overall computational power of R-CNN. In yet another novel work in the domain of disease prediction, authors have implemented machine learning techniques for the prediction of Dengue [27]. From the results it is concluded that LogitBoost ensemble technique is the most accurate one with accuracy equals to 92%. In the work [28] analysis on diabetes data is performed for the comparison of various machine learning techniques. In another significant study in the domain of disease prediction, authors are forecasting the

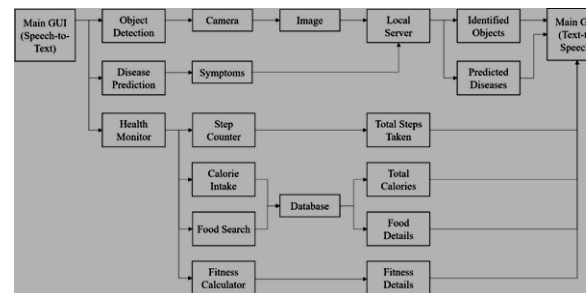


Figure 2: Control flow of the proposed android application.

casual effect association between Coronary Obstructive Pulmonary Disease and Cardiovascular Diseases [29].

3 System architecture

Visual impairment may be the result of an injury, disease, or some other condition. Visually impaired face problems in self-navigating outside known environments. They need to memorize details about their home environment. Furniture and other large obstacles must remain in one location to prevent injury. Individuals with low vision may find browsing websites problematic due to small fonts etc. They might also need to enlarge a screen significantly. Operating gadgets like mobile phones and tablets is also a challenge. The life of a visually impaired person can be transformed from a dependent individual to a productive and functional member of society who can read and write, use mobile phones, computers, and other gadgets efficiently with the help of modern assistive technologies.

Today, different solutions and technologies are present which have the potential to bring substantial change and improvement in the lives of people with visual impairment, especially the aging population. However, they have limited outreach, and end-users cannot fully realize their benefits. They must acquire essential awareness and know-how of using these technologies, as well as have the resources for obtaining them. Moreover, amidst abundant health management applications available today, applications catering to the visually impaired section of society are absent.

So, the need of the hour is to design affordable and accessible solutions to ease and significantly improve the daily routine and health management of the visually impaired. The main objective is to develop an application that would assist significantly in facilitating an independent & healthy lifestyle for the visually impaired.

- **Object Recognition** - A real-time object recognition module via the device camera is proposed. Objects will be identified and communicated to the user when they are in the camera's field of vision. Applying the concepts of computer vision, a deep learning-based architecture (Feature Pyramid Network with region Proposal Network) is used to identify and classify objects from the incoming stream of images.
- **Disease Prediction** - The user will be able to tell their symptoms in case he/she feels unwell and the same

will be checked against the data of various diseases. A graph database is implemented on Neo4j. The voice input of the user is queried and matched against the database to return one or multiple symptoms.

- **Monitoring of Daily Health Routine** - Daily health routine is monitored by calculating steps are taken, calorie intake, BMI, body fat percentage, and basal metabolic rate. Calorie intake for various food items is available to plan diet accordingly.
- **Voice Assistant** - For in-app interaction and navigation, a voice assistant is available. Google Speech-to-Text API is used to develop a text-to-speech module to communicate to the user and a speech-to-text module with natural language processing to process the user's input.

Figure 2 illustrates the flow of control of the android application. Initially, the user can select from the three modules - Object Recognition, Disease Prediction, and Health Monitor. Object Recognition is followed by capturing an image, which is sent to the local server hosting the deep learning model and then the identified objects are returned to the application and then to the user in the voice format. Disease Prediction is followed by taking voice input for symptoms, which are converted into text, sent to the local server for prediction of possible diseases, and returned to the user. A health monitor is comprised of keeping fitness records and the nutrition values of the food intake by the user.

4 Methodology

The main objective of the proposed health-based framework is to develop an android based app for visually impaired patients. The main modules of the proposed healthcare management tool for visually impaired patients are: 1) Object Detection, 2) Disease Prediction, 3) Real-time Monitoring of health Issues and 4) Voice assistant for In-App communication and Navigation. All these modules along with Dataset description are discussed in detail in this section.

4.1 Dataset description

A disease-symptom dataset was gathered from “WebMD” [30] and “MedicineNet”[31]. WebMD, founded in 1996, is one of the most-visited healthcare websites and contains data about various diseases, their corresponding symptoms, drug information, etc. MedicineNet is another site owned and operated by the WebMD consumer network. US Board-certified physicians and healthcare professionals maintain up-to-date information regarding diseases, symptoms, drugs, and remedies to the general masses in an easily understandable language. A graph database is created from the crawled data and the graph contains 149 disease nodes, 404 symptoms nodes, and 2126 “may cause” relationships i.e. un weighted directed edges from symptoms to corresponding disease. Also, the model was trained till 33 epochs on the Microsoft Common Objects

in Context dataset [32] which contains 330 K images, 1.5 million object instances and 91 object categories.

4.2 Disease prediction

The user will be able to tell their symptoms in case he/she feels unwell and same will be checked against the data of various diseases. A graph database is implemented in Neo4j. Neo4j is a graph database management system that contains nodes with directed edges or relationships between them. These nodes and edges can have labels and any number of attributes. Labels can be used to narrow searches. Querying is done using Cypher Query Language.

Neo4j is scalable and supports replication. It also supports ACID (Atomicity, Consistency, Isolation, and Durability) rules. Atomicity means the database considers all transaction operations as one whole unit or atom. Consistency guarantees that a transaction never leaves your database in a half-finished state. Isolation keeps transactions separated from each other until they're finished. Durability guarantees that the database will keep track of pending changes in such a way that the server can recover from an abnormal termination.

A graph database on Neo4j was implemented on the dataset to eliminate sparsity. The graph contains 149 disease nodes, 404 symptom nodes, and 2126 “(:Symptom)-[:MAY_CAUSE]->(:Disease)” relationships i.e. unweighted directed edges from symptoms to the corresponding disease. Weighted relationships between symptoms and diseases could not be obtained due to the unavailability of proprietary datasets.

The Neo4j database was linked with Python using Py2neo driver and queried using Cypher Query Language (CQL). Symptoms spoken by the user were matched with corresponding diseases and returned. The Python code was hosted on a Flask server to send information back and forth between the android application and the database.

4.3 Object recognition

The proposed model consists of a 51-layer Residual Network (ResNet-51) as the backbone for the Feature Pyramid Network (FPN). All 4 convolutional blocks of the Residual Network are used as a base for FPN. It has multiple prediction and upsampling layers. It has a lateral connection between the bottom-up pyramid and the top-down pyramid. It applies a 1x1 convolution layer before adding each layer. A 3x3 Conv layer is then applied and the result is used as a feature map by upper layers.

The pyramid of convolutional activation maps generated by FPN is passed to Region Proposal Network (RPN), eliminating the bottleneck of hardcoded algorithms like EdgeBoxes to get region proposals. RPN works as a first pass on the image and makes the binary decision if a region contains an object or not. It also outputs the confidence it has over the proposal.

The proposed regions are sent to the RoI Align layer which maps the proposed regions in the image to the convolutional features maps of FPN. RoI Align layer, unlike RoI Pooling layers, does not quantize the input



Figure 4: Relationships of “hypertensive disease” disease node and “pain chest” symptom node.

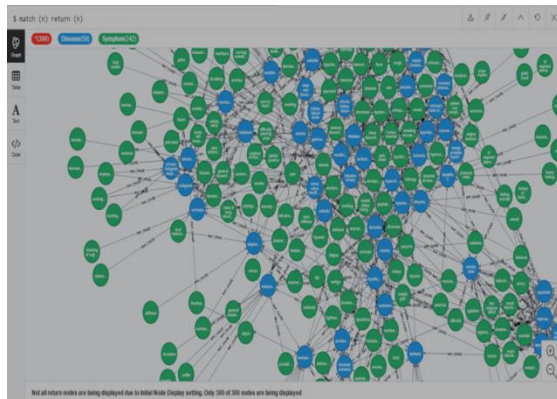


Figure 5: Neo4j graph database containing diseases, symptoms and their relationships.

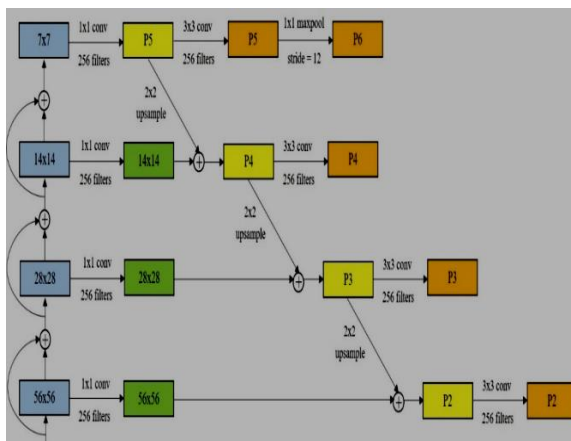


Figure 6: Feature pyramid network.

space by not rounding off to floor value i.e. it uses $a/16$ operation instead of $\lfloor a/16 \rfloor$ operation. This fixes the location misalignment issue.

The feature maps of candidate regions are then used by classification, bounding box, and mask predicting heads. Time distributed Keras layer is used to pass every feature map of the FPN to the heads. The entire heads share the first 2 fully connected layers which are implemented as convolutional layers; the first one has a dropout of 0.5 for regularization. The second convolutional layer has average pooling across the depth of the activation map making it function as fully connected layers.

Classification head: contains a fully connected layer that outputs logits for the proposal belonging to every

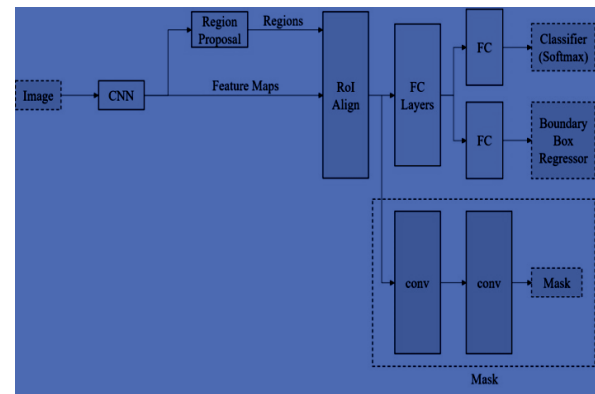


Figure 3: Architecture of object recognition model.

class or background class. The number of outputs is equal to a number of classes + 1 (background class). The softmax layer is used to calculate unnormalized log probabilities of the input belonging to every class and background class.

Bounding Box Regressor Head: Outputs bounding box deltas over the proposed bounding box by RPN. It consists of fully connected layers.

Mask Predicting head: The mask branch generates a mask of dimension 24×24 for each class for each proposed region. The total output is of size (number of classes + 1) * number of proposals. As the model tries to learn a mask for each class, masks are generated for every class. A mask is just a 24×24 binary grid which signifies the presence and absence of object instance for that pixel. It is implemented as 4 convolutional layers with batch normalization and ReLU activation followed by a fractionally stridden convolutional layer that upsamples the input. That is fed to another convolutional layer which has sigmoid activation to squash the input to -1 to 1. Loss is calculated using the mask associated with the ground truth class.

Monitoring of Daily Health Routine: Daily health routine is monitored by calculating steps are taken, calorie intake, BMI, body fat percentage, and basal metabolic rate. Calorie intake for various food items is available to plan diet accordingly. The Body Mass Index (BMI) is calculated using the height and weight of a person. It is defined as the body mass divided by the square of the body height and is universally expressed in units of kg/m^2 , resulting from mass in kilograms and height in meters. The value of BMI is used to categorize individuals as underweight, normal weight, or overweight. The body fat percentage (BFP) of an individual is the total mass of body fat divided by total body mass, multiplied by 100; body fat includes essential body fat and storage body fat. The body fat percentage is a measure of fitness level since it is the only body measurement that directly calculates a person's relative body composition without regard to height or weight. Metabolism comprises the processes that the body needs to function. Basal metabolic rate (BMR) is the amount of energy per unit of time that a person needs to keep the body functioning at rest.

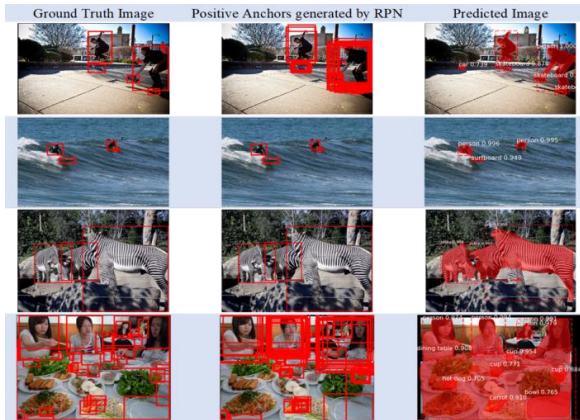


Figure 7: Results of object recognition model.

Voice Assistant: For in-app interaction and navigation, a voice assistant is available. The voice command that is input by the user in the English language is converted into an expression using Google Speech-to-Text API. WordNet is a large lexical database of the English language. Natural Language Toolkit is used to implement Part-of-Speech tagging on the expression. The similarities of the expression with the WordNet database are compared and the resulting command is executed or results of the operation are conveyed to the user using Google Text-to-Speech API.

5 Results and findings

The object recognition model was trained to 33 epochs with 1000 steps each epoch and 50 validation steps on the Microsoft Common Objects in Context dataset which contains 330K images, 1.5 million object instances, and 91 object categories. Mean image subtraction was done to center align data. The system used for training was expanded by an NVidia 1050 Ti OC Graphic Card for higher processing capability.

Object Recognition Model Hyper parameters

- No of epochs trained = 33
- Steps per epoch = 1000
- Validation steps = 50
- Threshold ratio of positive to negative RoI for training = 33
- Minimum probability value to accept a detected instance = 0.7
- Non-Maximum Suppression threshold = 0.3s
- Optimiser: Stochastic Gradient Descent with Momentum
 - Learning Rate = 0.002
 - Learning Momentum = 0.9
 - Weight Decay Regularization Strength = 0.0001

Mean Average Precision (MAP) is the standard single-number performance measure for comparing search algorithms. The MAP (mean Average Precision) score for the proposed model (ResNet-51-FPN) was 51.8%. Generation of feature maps in the proposed model using Feature Pyramid Networks is evaluated against existing models.

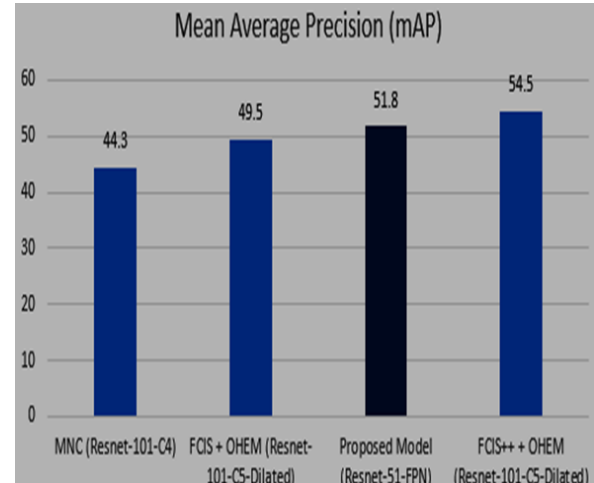


Figure 8: Performance comparison of proposed model with other object recognition models using mean average precision (mAP) score.

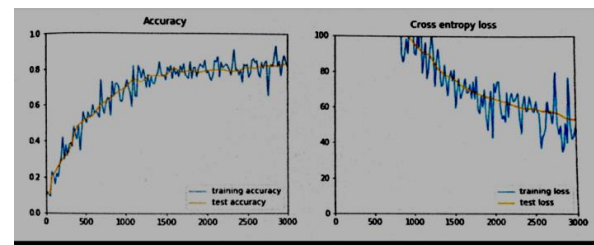


Figure 9: ANN-accuracy and cross entropy loss.

- Multi-task Network Cascades (MNC) [Dai, J., et al. 2015] consists of three networks, respectively differentiating instances, estimating masks, and categorizing objects. These networks form a cascaded structure and are designed to share their convolutional features.
- Fully Convolutional Instance-aware Semantic Segmentation (FCIS)[Li, Y., et al. 2016] proposed semantic segmentation and instance mask proposal. It detects and segments the object instances jointly and simultaneously.

The use of Feature Pyramid Networks to generate feature maps in the model fared better than existing models like MNC and FCIS+OHEM but worse than FCIS++OHEM due to the lesser number of layers in the Residual Network backbone of the architecture.

5.1 Comparative Analysis between ANN and CNN

In our experiments, ANN of 4 hidden layers of 20 neurons each with Batch Normalization gives 93% accurate results. Adding dropout technique in the model reduces the efficiency as it works for network which can afford to lose neurons. On the other hand, CNN even without over fitting countermeasures achieves accuracy of 98%, and with Batch Normalization and Dropout applied it gives accuracy approximately equals to 99.3%. CNN have repetitive blocks of neurons that are applied across space. At the training time, the weight gradients

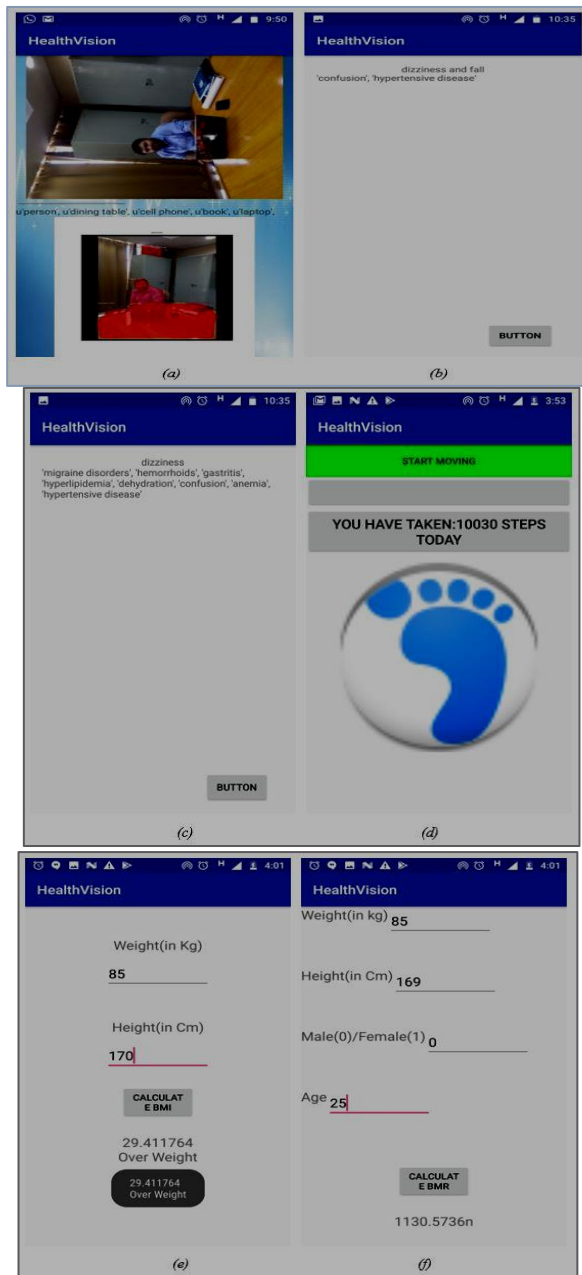


Figure 10: voice automated android application - (a) real-time object recognition (b) disease prediction when symptoms is dizziness (c) disease prediction when symptoms are dizziness and fall (d) no of steps taken daily (e) calculation of BMI (f) calculation

learned over various image patches are averaged which exploits the spatial or temporal invariance in object recognition.

For object recognition, the use of Feature Pyramid Network (FPN) provide valid proposals for most of the objects in the image while without FPN fails to provide valid proposals and misses out on a majority of objects. The model which has been trained till 40 epochs accurately captures even the small ground truth objects. That is why, FPN is used to generate feature maps for our recognition model. The test accuracy and test losses are visualized in the figures shown below.

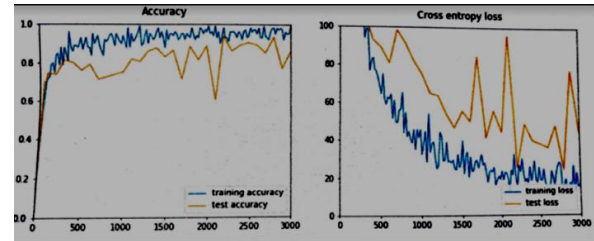


Figure 11: ANN + batch normalization-accuracy and cross entropy loss.

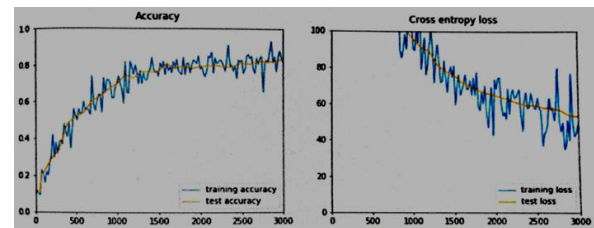


Figure 12: ANN + batch normalization+ dropout-accuracy and cross entropy loss.

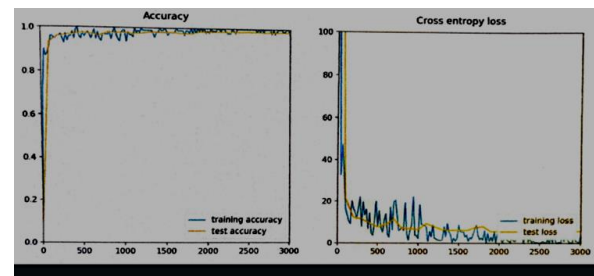


Figure 13: ANN- accuracy and cross entropy loss.

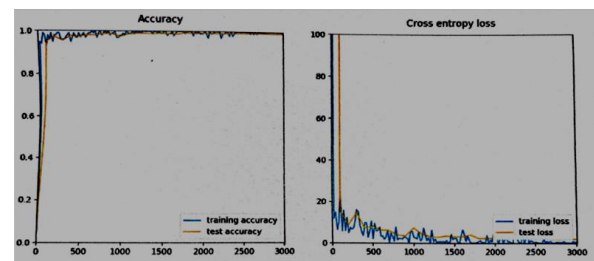


Figure 14: CNN + batch normalization- accuracy and cross entropy loss.

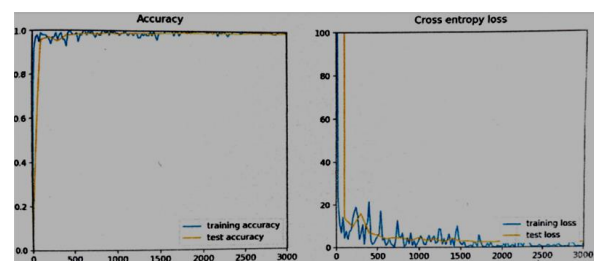


Figure 15: CNN + batch normalization+ dropout-accuracy and cross entropy loss.

5.2 Android interface

An Android application was developed for visually impaired with the following modules object recognition,

disease prediction, real time health monitoring, and voice assistant. Figure 10 (a)-(f) presents the screenshots of developed application interface.

6 Conclusion

The proposed application can open a new door to the possibilities of creating affordable and accessible solutions for improving the lifestyle and healthcare management of the visually impaired. The life of a visually impaired person can be transformed from a dependent individual to a productive and functional member of the society who is able to use mobile phones and other gadgets efficiently, detect objects around him and track his health with the help of such assistive technologies. A similar approach can also be extended to people suffering from other types of impairment.

The android application was successfully implemented with the following modules: object recognition, disease prediction, health monitor, and voice assistant. For object recognition, the use of Feature Pyramid Networks to generate feature maps in the model fared better than existing models like MNC and FCIS+OHEM but worse than FCIS+++OHEM due to the lesser number of layers in the Residual Network backbone. The graph database implementation of the disease-symptoms dataset gave the required results. However, graph algorithms could not be efficiently applied due to a lack of weighted edges i.e. probability that a symptom will cause a disease. This was due to the unavailability of appropriate non-proprietary datasets for general diseases.

7 Future research directions

The authors aim to improve the performance of the object recognition model by implementing aggregated residual transformations i.e. ResNeXt instead of ResNet for the backbone of Feature Pyramid Network. This model can also be extended for real-time depth realization to calculate the distance of the objects recognized. They also intend to enhance the disease prediction model by obtaining a more suitable weighted disease-symptom dataset, applying various graph algorithms, and adding drug suggestions. Lastly, they propose to upgrade the solution for healthcare management for the visually impaired from a voice-automated android application to a voice-automated hand-held device and add features like automatically alerting selected contacts in case of emergency.

References

- [1] Jonas et al., "Visual Impairment and Blindness Due to Macular Diseases Globally: A Systematic Review and Meta-Analysis", *American Journal of Ophthalmology*, vol. 158, no. 4, pp. 808-815, 2014. DOI: 10.1016/j.ajo.2014.06.012.
- [2] A. Gordoys et al., "An estimation of the worldwide economic and health burden of visual impairment," *Glob. Public Health*, vol. 7, no. 5, pp. 465–481, 2012. DOI: 10.1080/17441692.2011.634815
- [3] Bradski, G. R. "Computer vision face tracking for use in a perceptual user interface", 1998.
- [4] R. Velázquez, "Wearable Assistive Devices for the Blind," in *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment*, Springer, 2010, pp. 331–349.
- [5] [A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012, [Online].DOI : <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>.
- [6] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Apr. 2015, doi: 10.1007/s11263-015-0816-y.
- [7] C. Szegedy, A. Toshev, and D. Erhan, "Deep Neural Networks for Object Detection," *Neural Information Processing Systems*, 2013. <https://proceedings.neurips.cc/paper/2013/hash/f7cade80b7cc92b991cf4d2806d6bd78-Abstract.html> (accessed Mar. 11, 2022).
- [8] S.A.S, "Intelligent Heart Disease Prediction System Using Data Mining Techniques," *International Journal of healthcare & biomedical Research*, Volume: 1, Issue: 3, April 2013, pp. 94-101. DOI: <https://ijhbr.com/pdf/94-101.pdf>.
- [9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627-1645, Sept. 2010, doi: 10.1109/TPAMI.2009.167.
- [10] Kanuganti, S., Chang, Y., & Bock, L. U.S. Patent No. 9,836,996. Washington, DC: U.S. Patent and Trademark Office, 2017.
- [11] H. Fernandes, P. Costa, V. Filipe, L. Hadjileontiadis and J. Barroso, "Stereo vision in blind navigation assistance," *World Automation Congress*, 2010, pp. 1-6.
- [12] Milne LR, Bennett CL, Ladner RE. "The accessibility of mobile health sensors for blind users". In *International Technology and Persons with Disabilities Conference Scientific/Research Proceedings (CSUN 2014)*, Dec 2014 ,pp. 166-175.
- [13] Y. Tian, X. Yang, and A. Arditi, "Computer Vision-Based Door Detection for Accessibility of Unfamiliar Environments to Blind Persons," *Lecture Notes in Computer Science*, pp. 263–270, 2010, doi: 10.1007/978-3-642-14100-3_39.
- [14] B. Ando, C. O. Lombardo, and V. Marletta, "Smart homecare technologies for the visually impaired,"

- recent advances,” *Smart Homecare Technology and TeleHealth*, p. 9, Dec. 2014, doi:10.2147/shtts.56167.
- [15] Y. Ren, R. Werner, N. Pazzi and A. Boukerche, “Monitoring patients via a secure and mobile healthcare system,” in *IEEE Wireless Communications*, vol. 17, no. 1, pp. 59-65, February 2010, doi: 10.1109/MWC.2010.5416351.
- [16] Sachdeva N, Suomi R. Assistive technology for totally blind—barriers to adoption. *SOURCE IRIS: Selected Papers of the Information Systems Research Semina*. 2013;47.
- [17] M. H. S. Segler, T. Kogej, C. Tyrchan, and M. P. Waller, “Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks,” *ACS Central Science*, vol. 4, no. 1, pp. 120–131, Dec. 2017, doi: 10.1021/acscentsci.7b00512.
- [18] D. Dakopoulos and N. G. Bourbakis, “Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey,” in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 25-35, Jan. 2010, doi: 10.1109/TSMCC.2009.2021255.
- [19] Wexler, Y., & Shashua, A.. U.S. Patent No. 9,025,016. Washington, DC: U.S. Patent and Trademark Office, 2015.
- [20] Lewis, C. W., Mathers, D. R., Hilkes, R. G., Munger, R. J., & Colbeck, R. P.. U.S. Patent No. 8,135,227. Washington, DC: U.S. Patent and Trademark Office, 2012.
- [21] He K, Zhang X, Ren S, Sun J.” Deep residual learning for image recognition”. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778).
- [22] Girshick R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.
- [23] R. Girshick, J. Donahue, T. Darrell and J. Malik, “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142-158, 1 Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.
- [24] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *Neural Information Processing Systems*, 2015.
- [25] T.-Y. Lin et al., “Microsoft COCO: Common Objects in Context,” *Computer Vision – ECCV 2014*, pp. 740–755, 2014, doi: 10.1007/978-3-319-10602-1_48.
- [26] A. Chaudhuri, “Hierarchical Modified Fast R-CNN for Object Detection,” *Informatica*, vol. 45, no. 7, Dec. 2021, doi: 10.31449/inf.v45i7.3732.
- [27] N. Iqbal and M. Islam, “Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers,” *Informatica*, vol. 43, no. 3, Sep. 2019, doi: 10.31449/inf.v43i3.1548.
- [28] A. A. Abaker and F. A. Saeed, “A Comparative Analysis of Machine Learning Algorithms to Build a Predictive Model for Detecting Diabetes Complications,” *Informatica*, vol. 45, no. 1, Mar. 2021, doi: 10.31449/inf.v45i1.3111.
- [29] D. Panda, S. R. Dash, R. Ray, and S. Parida, “Predicting the Causal Effect Relationship Between COPD and Cardio Vascular Diseases,” *Informatica*, vol. 44, no. 4, Dec. 2020, doi:10.31449/inf.v44i4.3088.
- [30] WebMD, L. L. C. (2010). WebMD.
- [31] Shiel Jr, W. C. (2009). MedicineNet. com.
- [32] <http://www.har-dataset.org/>