# Leveraging CNN and Transfer Learning for Visionbased Human Activity Recognition

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Engineering degree in Computer Science and Engineering

By

K Rajeev (Reg.No - 39110499) K Kousik (Reg.No - 39110532)



# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING SCHOOL OF COMPUTING

# **SATHYABAMA**

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHISALAI,
CHENNAI - 600119

**NOVEMBER - 2022** 



# <u>SATHYABAMA</u>

INSTITUTE OF SCIENCE AND TECHNOLOGY



#### (DEEMED TO BE UNIVERSITY)

Accredited with —All grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

#### DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

# **BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the bonafide work of **K Rajeev(39110499)** and **K Kousik(39110532)** who carried out the Project Phase-1 entitled "Leveraging CNN and Transfer Learning for Vision-based Human Activity Recognition" under my supervision from June 2022 to November 2022.

Internal Guide
G. Anbuselvi M.Tech.(Ph.D)

**Head of the Department** 

Dr. L. LAKSHMANAN, M.E., Ph.D.

Submitted for Viva voce Examination held on\_\_\_\_\_

InternalExaminer

ExternalExaminer

#### **DECLARATION**

I, K Rajeev(Reg.No- 39110499), hereby declare that the Project Phase-1 Report entitled Leveraging CNN and Transfer Learning for Vision-based Human Activity Recognition" done by me under the guidance of MS.Anbu Selvi M.Tech.(Ph.D) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE: 27-10-22 PLACE:Chennai

SIGNATURE OF THECANDIDATE

Rajeor

#### **ACKNOWLEDGEMENT**

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph. D**, **Dean**, School of Computing, **Dr. L. Lakshmanan M.E., Ph.D.**, Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **MS.Anbu Selvi,** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my phase-1 project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

#### **ABSTRACT**

With the advent of the Internet of Things (IoT), there have been significant advancements in the area of human activity recognition (HAR) in recent years. HAR is applicable to wider application such as elderly care, anomalous behaviour detection and surveillance system. Several machine learning algorithms have been employed to predict the activities performed by the human in an environment. However, traditional machine learning approaches have been outperformed by feature engineering methods which can select an optimal set of features. On the contrary, it is known that deep learning models such as Convolutional Neural Networks (CNN) can extract features and reduce the computational cost automatically. In this paper, we use CNN model to predict human activities from Image Dataset model. Specifically, we employ transfer learning to get deep image features and trained machine learning classifiers. Our experimental results showed the accuracy of 96.95% using VGG-16. Our experimental results also confirmed the high performance of VGG-16 as compared to rest of the applied CNN models.

S.No	Topic	P-No
1	Abstract	5
2	Literature Review	7
3	Existing System	9
	3.1 Disadvantages of existing system	
4	Proposed System	9
	4.1 Advantages of proposed system	
5	Algorithm	10
6	System Architecture	11
7	Modules	12
	7.1 User	
	7.2 HAR System	
	7.3 VGG16	
	7.4 Transfer Learning	
8	System Requirements	13
	8.1 Hardware Requirements	
	8.2 Soft Requirements	

9	References	14

# **Literature Review:**

AUTHOR	YEAR OF PUBLICATION	DESCRIPTIONS	PROS	CONS	
R. Alazrai	2020	The complete deep learning model for Recognizing Human-to Human Interactions	The developed end to end deep learning model provides 86.3 accuracy for all humanto-human interactions recognition	The proposed model is not developed for group-to-group interactions. This model will work only for Human-to Human Interactions.	
T. Dobhal	2015	Binary Motion Image Deep learning	Binary Motion Image Deep learning model gives good accuracy for both 2D and 3D datasets consistent speed of action performed by a human	The model does not give a reasonable detection rate if more than one person is involved in the 3D image.	
M.Z. Uddin	2018	Kernel principal component analysis	KPCA outperform Support Vector Machine (SVM) and Artificial Neural Network (ANN)	It provides less accuracy for the real-time data.	
R. Janarthanan	2020	Unsupervised deep learning assisted reconstructed coder in the on-nodule wearable sensor for HAR	improves the feature selection and extraction using an unsupervised deep learning model	The performances degrade in large datasets with different types of human activities.	

AUTHOR	YEAR OF PUBLICATION	DESCRIPTIONS	PROS	CON
A. Jeyanthi Suresh	2020	Inception ResNet deep transfer learning method for HAR using LSTM	It provides the best accuracy score of 92 per cent and 91 per cent for the different data sets.	It takes a tre amount of tim
Y. Jia	2020	Multi-domain HAR based on Stepped-Frequency Continuous-wave radar using deep learning	Developed deep learning model increases the recognition accuracy by 1.3% by additionally introducing the range maps	The propose not develo group-to- interac
N. Zehra	2021	HAR using Ensemble Learning of Multiple CNN	It takes less amount of preprocessing time because the proposed model support automatic feature extractions	Model is no for concurre recogn
S. Ullah	2021	Sparse Feature Learning for Human Activity Recognition	It provides long term dependencies	It provid accuracy for time d

# **EXISTING SYSTEM:**

In the existingwork with wearable based or non-wearable based. Wearable based HAR system make use of wearable sensors that are attached on the human body. Wearable based HAR system are intrusive in nature. Non-wearable based HAR system do not require any sensors to attach on the human or to carry any device for activity recognition. Non-wearable based approach can be further categorised into sensor based HAR systems . Sensor based technology use RF signals from sensors, such as RFID, PIR sensors and Wifi signals to detect human activities. Sensor based HAR system are non-intrusive in nature but may not provide high accuracy.

#### **DISADVANTAGES OF EXISTING SYSTEM:**

- Require the optical sensors to be attached on the human and also demand the need of multiple camera settings.
- Wearable dives cost are high.
- Algorithm: Markerbased motion Capture (MoCap) Framework.

# **PROPOSED SYSTEM:**

The proposed System Vision based technology use videos, image frames from depth cameras or IR cameras to classify human activities. Video-based human activity recognition can be categorized as vision-based according to motion features. The vision based method make use of RGB or depth image. It does not require the user to carry any devices or to attach any sensors on the human. Therefore, this methodology is getting more consideration nowadays, consequently making the HAR framework simple and easy to be deployed in many applications. The most common type of deep learning method is Convolutional Neural Network (CNN). CNN are largely applied in areas related to computer vision. It consists series of convolution layers through which images are passed for processing.

#### **ADVANTAGES OF PROPOSED SYSTEM:**

- We use CNN to recognise human activities action recognition kinetics dataset.
- We use transfer learning to get deep image features and trained machine learning classifiers.
- Does not require the user to carry any devices or to attach any sensors on the human

**Algorithm**: Convolutional Neural Networks(CNN),VGG-16(alsocalled OxfordNet)

#### **Convolutional Neural Networks(CNN):**

A convolutional neural network (CNN or convnet) is a subset of machine learning. It is one of the various types of artificial neural networks which are used for different applications and data types. A CNN is a kind of network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data.

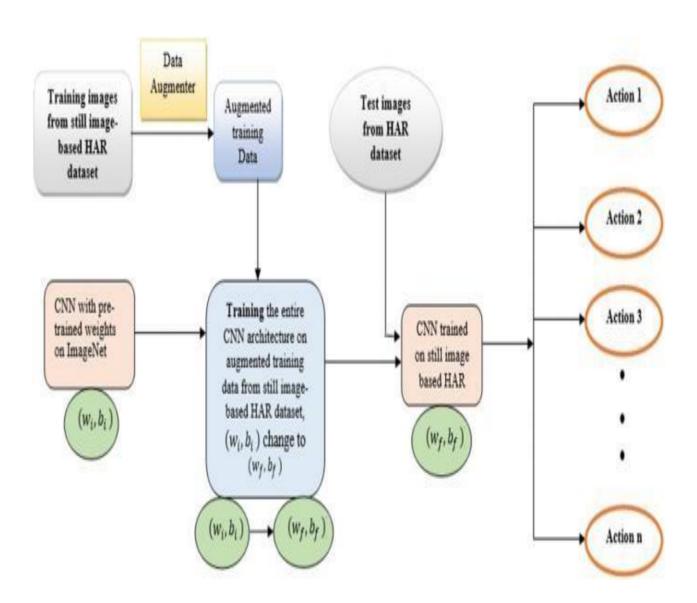
There are other types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice. This makes them highly suitable for computer vision (<u>CV</u>) tasks and for applications where object recognition is vital, such as self-driving cars and facial recognition.

# VGG-16(also called Oxford Net):

VGG16 proved to be a significant milestone in the quest of mankind to make computers "see" the world. A lot of effort has been put into improving this ability under the discipline of Computer Vision (CV) for a number of decades. VGG16 is one of the significant innovations that paved the way for several innovations that followed in this field.

It is a Convolutional Neural Network (CNN) model proposed by Karen Simonyan and Andrew Zisserman at the University of Oxford. The idea of the model was proposed in 2013, but the actual model was submitted during the ILSVRC ImageNet Challenge in 2014. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was an annual competition that evaluated algorithms for image classification (and object detection) at a large scale. They did well in the challenge but couldn't win.

# **SYSTEM ARCHITECTURE:**



#### **MODULES:**

- User
- HAR System
- VGG16
- Transfer Learning

#### **MODULES DESCRIPTION:**

#### User:

The User can start the project by running mainrun.py file. User has to give –input (Video file path). The open cv class VideoCapture(0) means primary camera of the system, VideoCapture(1) means secondary camera of the system. VideoCapture(Videfile path) means with out camera we can load the video file from the disk. Vgg16, Vgg19 has programitaically configured. User can change the model selection in the code and can run in multiple ways.

.

# **HAR System:**

Video-based human activity recognition can be categorised vision-based according. The vision based method make use of RGB or depth image. It does not require the user to carry any devices or to attach any sensors on the human. Therefore, this methodology is getting more consideration nowadays, consequently making the HAR framework simple and easy to be deployed in many applications. We first extracted the frames for each activities from the videos. Specifically, we use transfer learning to get deep image features and trained machine learning classifiers.

# **VGG16:**

VGG16 is a convolutional neural network model. Deep Convolutional Networks

for Large-Scale Image Recognition". The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous model submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using NVIDIA Titan Black GPU's.

# **Transfer Learning:**

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems. In this post, you will discover how you can use transfer learning to speed up training and improve the performance of your deep learning model.

# **SYSTEM REQUIREMENTS:**

# **HARDWARE REQUIREMENTS:**

> System : Intel Core i7.

➤ Hard Disk : 1TB.

Monitor : 15" LED

Input Devices : Keyboard, Mouse

➤ Ram : 16GB.

#### **SOFTWARE REQUIREMENTS**

> Operating system : Windows 10.

➤ Coding Language : Python

> Tool : Notepad++, Installer, GPS Signature

Database : SQLite

# **EXECUTION PART**

#### For each record it is provided

- - Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
- - Triaxial Angular velocity from the gyroscope.
- A 561-feature vector with time and frequency domain variables.
- - Its activity label.
- An identifier of the subject who carried out the experiment.

# The dataset includes the following files

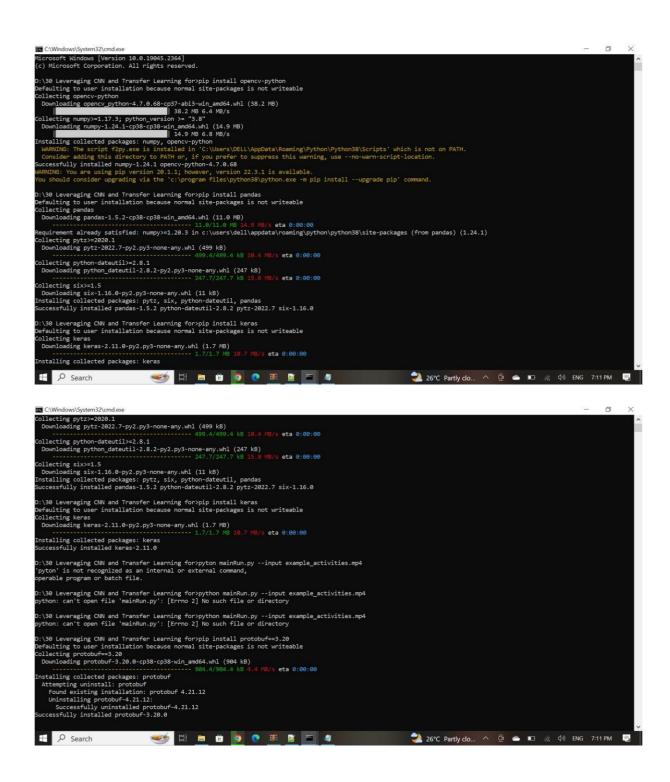
- 'README.txt'
- - 'features\_info.txt': Shows information about the variables used on the feature vector.
- - 'features.txt': List of all features.
- - 'activity\_labels.txt': Links the class labels with their activity name.
- - 'train/X\_train.txt': Training set.
- - 'train/y\_train.txt': Training labels.
- - 'test/X\_test.txt': Test set.
- - 'test/y\_test.txt': Test labels.

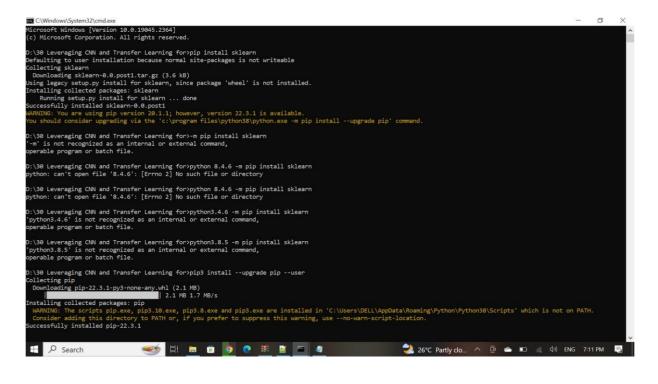
- The following files are available for the train and test data. Their descriptions are equivalent.
- - 'train/subject\_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
- - 'train/Inertial Signals/total\_acc\_x\_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128-element vector. The same description applies for the 'total\_acc\_x\_train.txt' and 'total\_acc\_z\_train.txt' files for the Y and Z axis.
- - 'train/Inertial Signals/body\_acc\_x\_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.
- - 'train/Inertial Signals/body\_gyro\_x\_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second.

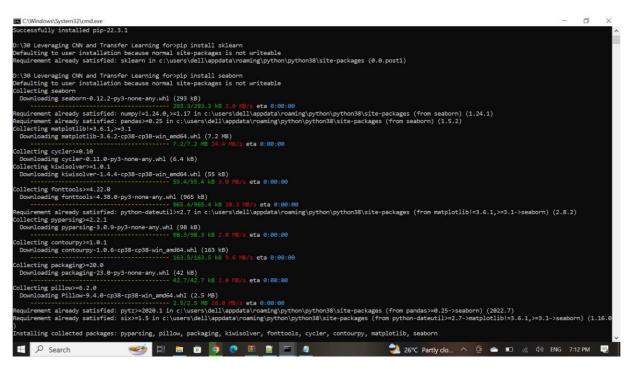
#### **CALCULATIONS USED FOR DATA SETS**

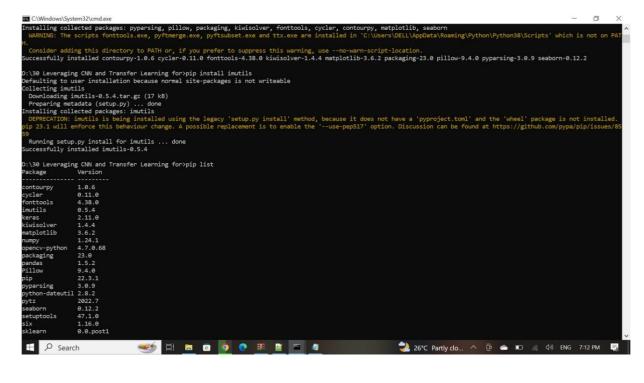
- Features are normalized and bounded within [-1,1].
- Each feature vector is a row on the text file.
- The units used for the accelerations (total and body) are 'g's (gravity of earth -> 9.80665 m/seg2).
- The gyroscope units are rad/seg.
- - A video of the experiment including an example of the 6 recorded activities with one of the participants can be seen in the following link: http://www.youtube.com/watch?v=XOEN9W05\_4A

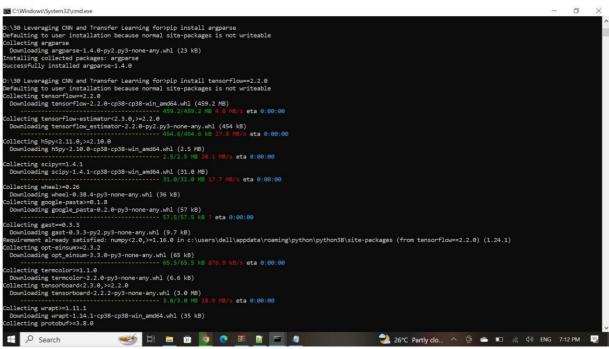
### **INSTALLATIONS OF PACKAGES**











```
Collecting protobufy=3.8.0

Downloading protobufy=3.8.12-cp38-cp38-win_amd64.wh1 (527 k8)

Collecting protobufy=3.8.6

Collecting protobufy=3.8.6

Collecting gracio>81.8.6

Downloading protobufy=3.8.6

Collecting gracio>81.8.6

Collecting arturpars=8.8.3

Downloading sturpars=1.6.3-cp32-cp38-win_amd64.wh1 (3.7 k8)

Collecting arturpars=8.8.3

Downloading sturpars=1.6.3-cp32-pg3-none-any.wh1 (128 k8)

Requirement already satisfied: six>1.12.8 in c:\user=\delta = 8:08:08

Requirement already satisfied: six>1.12.8 in c:\user=\delta = 8:08:08

Collecting keras-preprocessing=1.1.10

Downloading shol_py=0.7.0

Downloading keras-preprocessing=1.1.2-py2-py3-none-any.wh1 (128 k8)

Collecting keras-preprocessing=1.1.2-py2-py3-none-any.wh1 (42 k8)

Requirement already satisfied: setuptools=8.1.0.0 in c:\uper-cycapam files\python\8\text{Dython8\text{Ubin8\text{0}}}

Requirement already satisfied: setuptools=8.1.0.0 in c:\uper-cycapam files\python8\text{Ubin8\text{0}}

Collecting keras-preprocessing=1.1.1.2-py2-py3-none-any.wh1 (728 k8)

Downloading Narkdown-3.4.1-py3-none-any.wh1 (728 k8)

Collecting geogle-authc2,>=1.6.3

Downloading reposted suth-1.35.4-py2-py3-none-any.wh1 (128 k8)

Collecting geogle-authc2,>=1.6.3

Downloading requests-2.2.8.1-py3-none-any.wh1 (128 k8)

Collecting geogle-auth-1.35.4-py2-py3-none-any.wh1 (128 k8)

Collecting geogle-auth-2.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0.3-1.0
```

# **INSTALLATIONS DONE ON SYSTEM**

- Python-3.8.5-64-bit
- Python Launcher
- Notepad++ 8.4.6
- Notepad++ GPS Signature-8.4.6
- Notepad++ Installer-8.4.6
- Open CV Binaries in Python-4.5.5-cp39-cp39-win and 64.whl

# **INSTALLED PACKAGES**

•	absl-py	1.3.0
---	---------	-------

• astunparse 1.6.3

• cachetools 4.2.4

•	certifi	2022.12.7
•	charset-normalizer	2.1.1
•	contourpy	1.0.6
•	cycler	0.11.0
•	fonttools	4.38.0
•	gast	0.3.3
•	google-auth	1.35.0
•	google-auth-oauthlib	0.4.6
•	google-pasta	0.2.0
•	grpcio	1.51.1
•	h5py	2.10.0
•	idna	3.4
•	importlib-metadata	6.0.0
•	imutils	0.5.4
•	keras	2.11.0
•	Keras-Preprocessing	1.1.2
•	kiwisolver	1.4.4
•	Markdown	3.4.1
•	MarkupSafe	2.1.1
•	matplotlib	3.6.2
•	mysql-connector-python	8.0.32
•	numpy	1.24.1
•	oauthlib	3.2.2
•	opency-python	4.7.0.68

•	opt-einsum		3.3.0
•	packaging	23.0	
•	pandas	1.5.2	
•	Pillow		9.4.0
•	pip		22.3.1
•	protobuf	3.20.0	)
•	pyasn1	0.4.8	
•	pyasn1-modules		0.2.8
•	pyparsing	3.0.9	
•	python-dateutil		2.8.2
•	pytz		2022.7
•	requests	2.28.1	
•	requests-oauthlib		1.3.1
•	rsa		4.9
•	scipy		1.4.1
•	seaborn	0.12.2	2
•	setuptools	47.1.0	)
•	six		1.16.0
•	sklearn		0.0.post1A
•	tensorboard		2.2.2
•	tensorboard-plugin-wit		1.8.1
•	tensorflow	2.2.0	
•	tensorflow-estimator		2.2.0
•	termcolor	2.2.0	

• urllib3 1.26.13

• Werkzeug 2.2.2

• wheel 0.38.4

• wrapt 1.14.1

zipp 3.11.0

# **REFERENCE:**

- Samundra Deep, Xi Zheng Department of Computing Macquarie University
  Sydney, Australia Conference"Leveraging CNN and Transfer Learning for
  Vision-based Human Activity Recognition" International Accession
  Number:19572619.DOI:10.1109/ITNAC46935.2019.9078016.Telecommunication
  Networks and Applications Date Added to IEEE Xplore: 27 April 2020 INSPEC
- R. Alazrai, M. Hababeh, B.A. Alsaify, M.Z. Ali, M.I. Daoud, An End-to-End Deep Learning Framework for Recognizing Human-to-Human Interactions Using Wi-Fi Signals, IEEE Access. 8 (2020) 197695–197710.
- T. Dobhal, V. Shitole, G. Thomas, G. Navada, Human Activity Recognition using Binary Motion Image and Deep Learning, Procedia Comput. Sci. 58 (2015) 178–185.
- M.M. Hassan, M.Z. Uddin, A. Mohamed, A. Almogren, A robust human activity recognition system using smartphone sensors and deep learning, Futur. Gener. Comput. Syst. 81 (2018) 307–313.
- R. Janarthanan, S. Doss, S. Baskar, Optimized unsupervised deep learning assisted reconstructed coder in the on-nodule wearable sensor for human activity recognition, Meas. J. Int. Meas. Confed. 164 (2020) 108050.
- A. Jeyanthi Suresh, J. Visumathi, Inception ResNet deep transfer learning model for human action recognition using LSTM, Mater. Today Proc. (2020).
- Y. Jia, Y. Guo, G. Wang, R. Song, G. Cui, X. Zhong, Multi-frequency and multi-domain human activity recognition based on SFCW radar using deep learning, Neurocomputing.