Human Activity recognition using Surface EMG data and Deep Learning

Submitted in partial fulfilment of the requirements

of the degree of

Bachelor of Technology

by

Akella Lakshmana Sai Srikar	**204103**
Chinta Phani Rama Vaibhav	**204115**
Yagna Satwik	**204178**

Supervisor:

Dr. Mohammad Farukh Hashmi

Assistant Professor

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL

2023

Approval Sheet

This Project Work entitled **"Human Activity recognition using Surface EMG data and Deep Learning"** by Akella Lakshmana Sai Srikar (204103), Chinta Phani Rama Vaibhav (204115), Yagna Satwik (204178) is approved for the degree of Bachelor of Technology in Electronics and Communication Engineering.

Examiners

____

____

____

____

____

Supervisor

____

Chairman

____

Date: __________

Place: __________

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Roll No.	Name of the student	Signature of the student
204103	Akella Lakshmana Sai Srikar	
204115	Chinta Phani Rama Vaibhav	
204178	Yagna Satwik	

Date: _______________

Certificate

This is to certify that the Project work entitled **"Human Activity recognition using Surface EMG data and Deep Learning"** is a bonafide record of work carried out by "Akella Lakshmana Sai Srikar(204103), Chinta Phani Rama Vaibhav(204115), Yagna Satwik (204178)", submitted to the faculty of Department of Electronics and Communication Engineering, in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Electronics and Communication Engineering at National Institute of Technology, Warangal during the academic year 2023-24.

Dr. D. Vakula
Head of the Department,
Department of ECE
NIT Warangal

Dr. Mohammad Farukh Hashmi
Assistant Professor
Department of ECE

Acknowledgement

We would like to thank our supervisor, Dr. Mohammad Farukh Hashmi for providing us with invaluable guidance and support throughout the project. His expert knowledge and insightful suggestions have been instrumental in shaping the project's direction. Also, we would like to express our sincere gratitude and appreciation to all those who have contributed to this project on "**Human Activity recognition using Surface EMG data and Deep Learning".**

Abstract

Human Activity Recognition (HAR) is a technology which utilizes sensor data of any kind to recognize or categorize human actions. HAR is actively used in every field of work, for healthcare, safety, athletics, Sports and Security. The applications pertaining to this field and technology are limitless as are the challenges and current scope for research. This report aims to use a non-traditional approach to Sensor data collection and study and judge various possible Machine Learning and Deep Learning Techniques to find out the optimum and most effective model for activity recognition.

The project uses surface - EMG data to prepare and train its model since EMG data has few outliers and less noise compared to data like IMU and EEG. Two particular datasets are used in the process: The EMG Physical Action Data Set [1] and The EMG data in lower limb [2]. Initially the research explores various traditional machine learning approaches to this activity recognition challenge. Machine Learning Models such as Random Forest, CNN, LSTM and the CNN+LSTM Hybrid model gave an accuracy of 91%, while traditional models fell below the 90% mark. Further Time Frequency Analysis techniques are tested, first the STFT and moving further SS-STFT. The STFT+CNN model generated an impressive accuracy of 95%. Transfer Learning techniques and pre-trained models with fully connected layers were explored to leverage spatial data and improve performance. The MobileNet + CNN architecture proved to give an even higher accuracy at 97.34%. Moving further, the report also elaborates on the use of SS-STFT as an alternative to STFT along with the MobileNet model, which produced the maximum accuracy of 98.56%. The use of MobileNet had greatly reduced computational load and also produced greater accuracies and efficiency in the process.

List of Contents

List of Figures

List of Tables

List of Abbreviations

HAR: Human Activity Recognition

sEMG: Surface Electromyography

STFT: Short Time Fourier Transform

SST: Synchro Squeezing Transform

SS – STFT: Synchro Squeezing Short Time Fourier Transform

TFI: Time Frequency Images

UCI: University of California, Irvine

CSV: Comma Separated Values

ML: Machine Learning

DL: Deep Learning

TL: Transfer Learning

RF: Random Forest

CNN: Convolutional Neural Network

LSTM: Long Short Term Memory

ANN: Artificial Neural Network

RNN: Recurrent Neural Network

GRU: Gated Recurrent Unit

VGG: Visual Geometry Group (VGGNet, a convolutional neural network architecture)

PCA: Principal Component Analysis

Chapter 1

Introduction

1.1 Motivation

Human Activity Recognition (HAR) has emerged powerful for applications in healthcare, fitness monitoring, and human-computer interaction (HCI). Deep learning algorithms excel at extracting complex patterns from data [3], making them ideal for analyzing the intricate signals from sEMG sensors. We aim to develop a robust and accurate HAR model using deep learning techniques. This project hence, explores the potential of sEMG data for activity recognition by leveraging deep learning.

Figure 1.1: Various daily life recognizable Human Activities

1.2 Problem statement

Existing HAR systems often rely on sensors like accelerometers [4], which face limitations. These limitations include susceptibility to environmental factors (triggering false positives) and difficulty in differentiating between activities with similar movement patterns. This project addresses these limitations by exploring the use of surface Electromyography (sEMG) data for activity recognition. sEMG data offers a unique advantage as it directly captures the electrical activity within muscles during movement. However, analyzing this complex data for activity classification presents a challenge. Traditional machine learning (ML) algorithms might struggle to identify the intricate patterns within sEMG signals [5], leading to inaccurate activity recognition.

1.3 Objective of the report

The development and testing of a deep learning (DL) model using surface electromyography (sEMG) data for human activity recognition (HAR) is described in this report. The major goal is to establish accurate and efficient classification of different human activities. We look at numerous approaches, including standard machine learning and deep learning models such as Random Forest, CNN, and LSTM. We evaluate the usefulness of the Short-Time Fourier Transform (STFT) as a time frequency representation, as well as transfer learning. Additionally, we investigate Synchronous Squeezing STFT (SS-STFT) [6] as an alternative for STFT. Finally, we create and test 2D CNN architectures [7] that are specifically designed to process the time-frequency representations acquired using these methodologies. All attained results are tabulated and segmented to review and discuss the most efficient system for HAR with the present computational and cost effective methods.

1.4 Scope of the Report

This project focuses on creating a deep learning model for human activity recognition (HAR) utilizing surface electromyography (sEMG) data. The emphasis is on comparing several approaches to determine the most effective model architecture for accurate sEMG-based HAR. This report contains the following:

- **Model Architectures:** We examine the performance of classic ML and DL models such as Random Forest, CNNs, and LSTM networks for sEMG-based HAR. Furthermore, we investigate transfer learning using pre-trained models like VGG19 and ResNet50.
- **Time Frequency representation:** We investigate the efficiency of the Short-Time Fourier Transform (STFT) for converting raw sEMG data to a time-frequency representation. Furthermore, we investigate Synchronous Squeezing STFT (SS STFT) as an alternate pre-processing strategy that may increase feature extraction.
- **2D CNN Development:** We create and test 2D CNN architectures that are specifically designed to handle the time-frequency data received from STFT and SS STFT. These CNN models are created to learn key spatial and temporal properties from modified data in order to classify activities reliably.

1.5 Applications

Electromyography (EMG) has traditionally been used in medical settings to diagnose neuromuscular diseases. However, its uses are constantly increasing into numerous industries. In Human-Computer Interaction (HCI), EMG enables those with limitations [8].

With regard to this project, Electromyography (EMG) analyses electrical activity in muscles, offering valuable insights into muscle health. By analyzing the electrical signals, your models can extract features like muscle activation patterns and frequency content, allowing them to differentiate between activities based on their unique "electrical signatures." Compared to external motion sensors, sEMG data's detailed representation of muscle activity has the potential to significantly improve classification accuracy, especially for subtle movements or activities with similar external patterns.

1.6 Organization of the Report

This project looks at how to create a deep learning model for human activity recognition (HAR) using sEMG data. It walks the reader through the research procedure and the results.

The introduction discusses HAR limits and sEMG possibilities, followed by project objectives. The literature survey examines existing sEMG-based HAR research and places your idea within this discipline. The methods and materials section starts by giving the details of the datasets used, followed by a brief description of all the techniques used in our project.

The Methodology section delves into the approaches investigated. It describes the use of several machine learning models and time frequency representation approaches such as STFT and SS STFT and the advantages of each time frequency representation for feature extraction. Additionally, this section discusses transfer learning with pre-trained models and the design of 2D CNN architectures for handling time-frequency data. The results and discussion section shows the evaluation results of the different models and approaches. It contains all the findings, including model performance metrics and the effectiveness of pre-processing techniques and architectures. Finally, the conclusion summarizes key findings, reiterates achieved objectives, highlights the most effective model and pre-processing techniques, discusses limitations, and suggests future research directions. The report concludes with a complete list of references.

Chapter 2

Literature Survey

Title: Deep Learning of EMG Time–Frequency Representations for Identifying Normal and Aggressive Actions [9]

Authors & Publication: H. Alaskar, International Journal of computer science and Network Security, Vol. 18, No. 12 (2018)

This project explores using convolutional neural networks (CNNs) to automatically classify physical activities from EMG signals. Unlike traditional methods requiring manual feature engineering, this approach utilizes an "end-to-end" deep learning technique. The project investigates how different time-frequency representations (spectrogram vs. scalogram) of EMG data affect CNN performance.

Results show that scalogram representations lead to higher accuracy, with a simple CNN achieving 94.61% in classifying normal and aggressive activities. This suggests that CNNs with scalogram representations hold promise for accurate activity recognition using EMG data.

Title: Hybrid Deep Learning Approaches for sEMG Signal-Based Lower Limb Activity Recognition [10]

Authors & Publication: Ankit Vijayvargiya et al. Mathematical Problems in Engineering (Vol. 2022)

This paper delves into the innovative strategy of employing a hybrid deep learning model, specifically a combination of CNN and GRU. This model incorporates advanced techniques such as wavelet denoising and segmentation, resulting in an impressive accuracy of 98.61%.

However, a noteworthy limitation is the focus on a restricted set of activities, namely walking, sitting, and standing. To enhance the effectiveness of sarcopenia determination, a broader spectrum of activities is deemed essential.

Title: Surface EMG Signals and deep transfer learning-based physical action classification [11]

Authors & Publication: Demir et al. Neural Computing and Applications (Vol: 31) (2019)

This paper explores an advanced approach for human action classification through the utilization of surface electromyography (EMG) signals and transfer learning. The representation of these signals involves the creation of spectrograms using short-time Fourier transform.

Remarkably impressive accuracy of 99.04% is achieved by employing a combination of deep features from AlexNet's fc6 and fc7 layers, as well as VGG16's fc6 and fc7 layers, followed by support vector machine (SVM) classification. This paper is commendable not only for its high accuracy but also for the intricacy of its neural structure, marking a significant contribution to the field.

Title: Pattern Matching for Real-Time Extraction of Fast and Slow Spectral Components From sEMG Signals [12]

Authors & publication: Alvaro Costa-Garcia et al. IEEE Transactions on Neural Systems and Rehabilitation Engineering, volume 31 (2023)

This study investigates the possibilities of using surface electromyography (sEMG) spectrum decomposition to assess muscle performance, motor learning, and the effects of ageing on lower limb (LL) muscle activities under sit-stand-sit disruptions. The study suggests creating a system that can extract spectral components in real-time. This system would use a pre-defined library that is based on a huge dataset. A model library is designed to meet the specific requirements of a real-time application. It demonstrates a balanced performance in terms of sEMG reconstruction accuracy and component stability using 300 data segments The components of the model exhibit minimal variation between subjects, suggesting that a small number of subjects can accurately represent a wide range of physiological variables. Motion specific pre-processing stages are implemented to ensure accurate extraction of components for isotonic and isometric muscle contractions. The study shows how fast and slow components can be extracted from sEMG signals, suggesting the potential for real-time classification using multivariable linear regression. The system appears promising because it can extract spectral components from just 20 muscle activations, making it versatile for a wide range of daily movements.

Title: Human Activity Recognition from Body Sensor Data using Deep Learning [13]

Authors & Publication: Mohammad Mehedi Hassan et al. Journal of Medical Systems, Volume 42, article number 99, (2018)

This project tackles the challenge of accurately recognizing human activities using wearable sensor data. This is addressed by proposing a Deep Belief Network (DBN) model. The process involves first extracting relevant information from the raw sensor data. Following this, the extracted features undergo further refinement using Kernel Principal Component Analysis (KPCA) and Linear Discriminant Analysis (LDA). These techniques improve the robustness of the features and make them more suitable for fast and accurate activity recognition. Finally, the refined features are used to train the DBN model.

Experiments using real-world wearable sensor data demonstrate that the proposed DBN model outperforms other algorithms, achieving a high level of accuracy in activity recognition. This paves the way for utilizing DBNs in developing reliable and efficient wearable sensor systems for various healthcare applications.

Title: A Deep Learning-based Model for Human Activity Recognition using Biosensors embedded into a Smart Knee Bandage [14]

Authors & Publications: Sakorn Mekruksavanich et al. Procedia Computer Science (Vol. 214) (2022)

This research paper proposes a sophisticated model for human activity recognition. The model architecture comprises a combination of Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and ResNeXt networks. This amalgamation is designed to extract optimal features from each network individually, contributing to an integrated system that achieves an accuracy level of 91.60%.

Notably, the model is adept at classifying activities based on distinct biosensor signals, namely Electromyography (EMG), Electrogastrogram (EGM), and Inertial Measurement Unit (IMU) data. The utilization of such multimodal sensor information enhances the versatility of applications, establishing this model as an exemplary choice for human activity recognition in diverse scenarios.

Title: A CSI-Based Human Activity Recognition Using Deep Learning [15]

Authors & Publication: Fard Moshiri, P et al. Sensors 2021, 21, 7225.

This project explores a novel approach to human activity recognition (HAR) for elderly care using WiFi signals. Traditional methods like cameras or wearable sensors can be intrusive or raise privacy concerns. This project leverages WiFi's ubiquity to monitor daily activities.The system uses a Raspberry Pi to collect Channel State Information (CSI) data, a characteristic of WiFi signals, for different activities. This CSI data is then converted into images. These images are fed into a 2D CNN for activity classification.

The proposed method achieved an impressive accuracy of around 95% for recognizing seven daily activities. This outperforms other techniques like 1D-CNN, LSTM and Bi-directional LSTM, demonstrating the potential of WiFi-based CSI data and 2D CNNs for accurate and privacy-preserving HAR in elderly care applications.

Title: Human activity recognition with smartphone sensors using deep learning neural networks [16]

Authors & Publication: Charissa Ann Ronao et al. Expert Systems with Applications, Volume 59, 2016

This study looks into a deep learning approach for human activity recognition (HAR) using smartphone sensor data. The proposed approach uses a CNN to make use of the inherent characteristics of human behaviours and time-series data. This allows the CNN to automatically extract features from raw sensor data, eliminating the need for manual engineering. The study investigates the effectiveness of various network topologies, showing that CNNs learn more complicated properties with each layer. The best settings for identifying temporal links in data were also determined, with broader windows and lower pooling sizes performing better.

This approach achieved impressive results, particularly for recognizing similar activities that were previously challenging. Compared to other techniques, the CNN achieved a significant accuracy improvement, reaching 94.79% on raw sensor data and 95.75% when incorporating additional temporal information. These findings suggest that CNNs hold great promise for accurate and efficient HAR using smartphone sensors.

Title: Analysis of aging effect on lower limb muscle activity using short time Fourier transform and wavelet decomposition of electromyography signal [17]

Authors & Publication: Tawhidul Islam Khan et al. AIP Advances, Volume 13, Issue 5 (2023)

This paper explores how aging affects the muscles in our lower limbs during sit-stand-sit movements, especially for those with knee osteoarthritis (OA). Involving 84 participants, including both healthy individuals and knee OA patients, the study used surface electromyography (EMG) technology and various techniques like short time Fourier transform (STFT) and Wavelet Transform (WT) to analyze muscle activities. The findings show that middle and older age participants had higher muscle activities than younger individuals but lower than OA patients. Importantly, the research suggests that muscle weakening has a more significant impact on knee OA development than the natural aging process. The study emphasizes the importance of early detection of muscle deterioration for understanding OA progression and tailoring treatments for older individuals. For future work, the authors suggest a more extensive study with machine learning algorithms to better quantify the aging effects on muscle activities.

Title: Worker Activity Recognition in Smart Manufacturing Using IMU and sEMG Signals with Convolutional Neural Networks [18]

Authors & Publication: Wenjin Tao et al. Procedia Manufacturing, Volume 26, 2018,

This project investigates worker activity recognition in smart factories to improve performance evaluation and provide on-site augmented reality instructions. It utilizes a Myo armband equipped with Inertial Measurement Unit (IMU) and surface electromyography (sEMG) sensors. The IMU data is processed to create an "activity image" that reflects worker movements. This image is then fed into a CNN to extract high-level movement features. Simultaneously, sEMG signals are analyzed to determine muscle activation levels.

These features extracted from both IMU data and muscle activation are then combined and used to classify worker activities into six categories related to assembly tasks. The model achieved impressive accuracy (98% in a standard test and 87% in a leave-one-out test) on a dataset encompassing these common activities.

Title: sEMG signal filtering study using synchro-squeezing wavelet transform with differential evolution optimized threshold [19]

Authors & Publication: Chuanjiang Li et al. Results in Engineering, Volume 18, June 2023

This paper provides a comprehensive and systematic study on the feasibility and effectiveness of using Synchro-squeezing wavelet transform (SWT) to denoise sEMG signals, the proposed algorithm was used for gesture recognition but can be extended to other sEMG applications as well. The sEMG signals were mixed with three noises baseline drift, power line interference, and white gaussian noise. Then SWT was used for filtering and three metrics were used to evaluate the performance, SNR, RMS error, R-squared value, the algorithm achieved high denoising performance in comparison to classical Infinite Impulse Response (IIR) and the empirical mode decomposition (EMD) algorithm. The gesture recognition accuracy of the SWT denoised signal was 95.95%. The methods proposed in this paper can also be applied to EEG signals fields like Brain – machine interfaces, Intelligent prosthetics, to achieve precise prosthetic movement and control.

Title: Evaluation of Feature Extraction and Recognition for Activity Monitoring and Fall Detection Based on Wearable sEMG Sensors [20]

Authors & Publication: Xugang Xi et al. Sensors 2017, 17, 1229.

This study investigates using leg muscle EMG data to track everyday activities and detect falls in the elderly. The researchers analyzed 15 approaches for collecting characteristics from EMG data and evaluated 5 classification systems for accuracy and efficiency. The experiments entailed recording EMG data from three volunteers while they were conducting daily activities and a simulated fall. The investigation demonstrated that employing particular features such as "Wilson Amplitude" (WAMP), high accuracy and rapid processing may be achieved for both monitoring and fall detection. For activity monitoring, a "Gaussian Kernel Support Vector Machine" classifier obtained recognition rates of more than 96% using WAMP or another feature. Fall detection required a trade-off between accuracy and quickness. A different classifier provided the most accuracy, but it was also the slowest. A good balance was discovered with a "Gaussian Kernel Fisher Linear Discriminant Analysis" classifier employing WAMP, which achieved excellent fall detection accuracy (more than 98%) while processing quickly.

Title: sEMG-Based Gesture Recognition with Convolution Neural Networks [21]

Authors & Publication: Zhen Ding et al**.** Sustainability 2018, 10, 1865

Existing methods for recognizing limb movements using sEMG signals lose accuracy due to information loss during feature extraction. To overcome this, researchers propose a new deep learning architecture - parallel multiple-scale convolution. This approach uses larger filters than typical methods, capturing more data from the sEMG signal. Additionally, it considers the independence of different muscles, potentially improving motion distinction. Tested on a standard dataset, this architecture achieved the highest accuracy, suggesting that using larger filters, parallel processing, and considering muscle independence can significantly improve sEMG-based limb motion recognition.

Chapter 3

Methods and Materials

3.1 Details of the Datasets

The EMG Physical Action Data Set consists of 10 activities classified as normal and 10 actions classified as aggressive. The data was collected from four participants, utilizing the Delsys EMG wireless device. There are a total of 8 electrodes, each matching to one of the 8 input time series. Each electrode is associated with a certain muscle channel, labelled as ch1 upto ch8. Each time series consists of around 10,000 samples, which correspond to 15 acts performed by each subject in an experimental session. Crucially, the dataset is free of any missing values.

The EMG dataset in lower limb was obtained form 22 subjects. Participants were engaged in a variety of tasks, including: Legs bent up, legs extended from sitting position and walking. Data was compiled using his Data LOG (MWX8) from Biometrics Ltd. Four electrodes on the rectus femoris (RF), semitendinosus (ST), biceps femoris (BF), and vastus medialis (VM) along with a goniometer in the knee were used for the acquisition process. The goniometer values were not considered for training. This dataset also has no missing values.

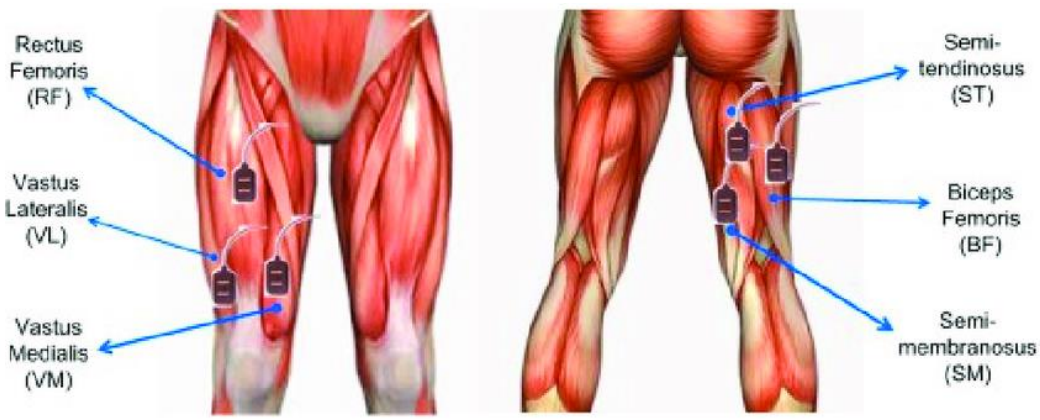


Figure 3.1 Position of muscles (VM, ST, BF, RF) [22]

3.2 Working of Random Forest & Deep learning Models

Random Forest:

It is a supervised machine learning algorithm effective for both regression and classification problems. It makes use of ensemble learning by combining several decision trees, each trained on different subsets of the dataset. Instead of relying on the prediction of a single tree, the random forest makes predictions based on the majority vote of all the trees [23]. Increasing the quantity of trees in the forest enhances accuracy and mitigates the issue of overfitting. Figure 3.2 illustrates the functioning of the Random Forest algorithm.

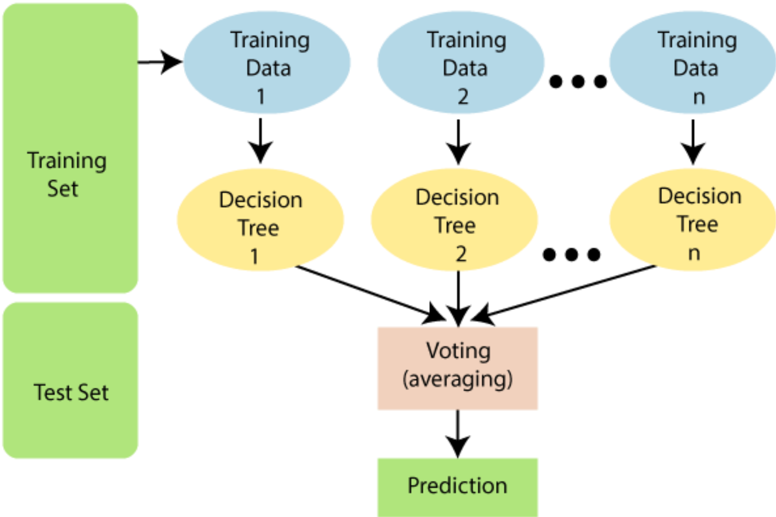


Figure 3.2: Working of Random Forest [24]

Convolutional Neural Networks (CNN):

A Convolutional Neural Network (CNN) is a type of deep learning architecture widely employed for tasks like image classification, though its versatility extends to other data types such as audio and text. It's an advanced iteration of the Artificial Neural Network (ANN), specifically tailored for data arranged in grid-like matrices.

Here's a breakdown of its key components, the same is shown in Figure 3.3:

1. **Input Layer:** This is where the raw data is fed into the network. In the case of images, each pixel's intensity values serve as input.

2. **Convolutional Layer:** These layers apply convolution operations to the input data, utilizing filters or kernels to extract features. These features can represent various patterns present in the data, like edges, textures, or shapes.

3. **Pooling Layer:** After convolution, in order to reduce the dimensionality of the feature maps without sacrificing the most important information, pooling layers down sample them. Max pooling and average pooling are two popular pooling techniques.

4. **Fully Connected Layers:** These layers link every neuron in the preceding layer to every neuron in the next layer. They make it possible for the network to understand complex relationships between high-level features retrieved by convolutional and pooling layers.

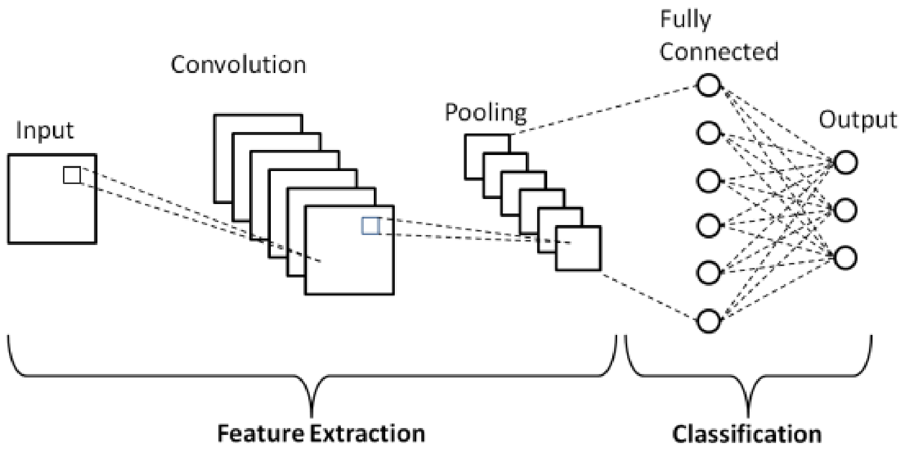


Figure 3.3: Working of CNN [25]

CNNs leverage the hierarchical arrangement of these layers to progressively learn representations of increasing complexity. Through repeated application of convolution and pooling operations, the network can effectively capture hierarchical patterns present in the data, enabling robust feature extraction and classification.

Long Short Term Memory (LSTM):

Long Short-Term Memory is a specific type of recurrent neural network (RNN) that is capable of capturing and retaining long-term relationships in sequential data, such as time series, text, and speech. Memory cells and gates are utilised to regulate the information flow, enabling the selective retention or discarding of information as required. This effectively circumvents the vanishing gradient problem commonly encountered in conventional RNNs. [26]. There are three different gates in an LSTM: input gate, forget gate, and output gate, sigmoid functions which produce an output between 0 and 1 are used to implement them.

- The input gate determines the specific information that will be stored in the memory cell. It is activated when the input is deemed significant and deactivated when it is not.

- The forget gate determines the information to be eliminated from the memory cell. It is activated when the information becomes irrelevant and vice versa.
- Determining which information to integrate into the LSTM's output is the duty of the output gate. It has been programmed to selectively reveal pertinent information and withhold it when it is not.

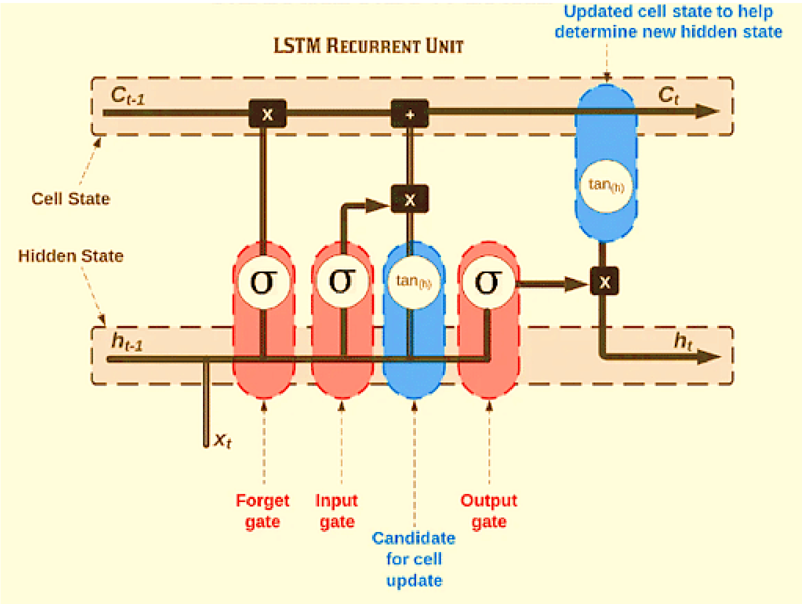


Figure 3.4: LSTM Recurrent unit [27]

The LSTM cell is as shown in Figure 3.4, has a memory cell that retains data from preceding time steps and employs that information to modulate the cell's output during the current time step. By transferring the output of each LSTM cell to the subsequent cell in the network, the LSTM is capable of analysing and processing sequential data across numerous time steps.

CNN-LSTM Hybrid:

It combines the strengths of CNN, LSTM networks for sequential data processing. In this architecture, CNNs are employed for feature extraction from input sequences, capturing spatial patterns and dependencies [28]. These extracted features are then fed into LSTM layers, which effectively model temporal dependencies and long-range dependencies in the sequence. This hybrid model is particularly beneficial for tasks involving sequential data with spatial structures, such as video classification, speech recognition, and time-series forecasting. By integrating CNNs and LSTMs, the model can leverage both local and global information, leading to improved performance in tasks requiring both spatial and temporal understanding.

3.3 Short Time Fourier Transform

The Short-Time Fourier Transform (STFT) is a signal processing technique employed to analyse the temporal evolution of frequency content within a signal. This method dissects a signal into brief segments, subjecting each segment to Fourier Transform. The outcome is a time-frequency representation, elucidating the variation of frequency components across distinct time intervals.

The generation of time-frequency images involves plotting the magnitude or power of frequency components acquired through STFT analysis against time and frequency axes. Conventionally, the x-axis signifies time, the y-axis signifies frequency and colour or intensity denotes the magnitude or power of frequency components. These visual representations, commonly referred to as spectrograms, visually articulate the dynamic changes in frequency content over time within the signal. Spectrograms, integral in analyzing non-stationary signals characterized by varying frequency components over time, find applications in diverse domains like audio signal processing, radar signal analysis, speech recognition, and vibration analysis. They contribute significantly to comprehending signal characteristics across both time and frequency domains.

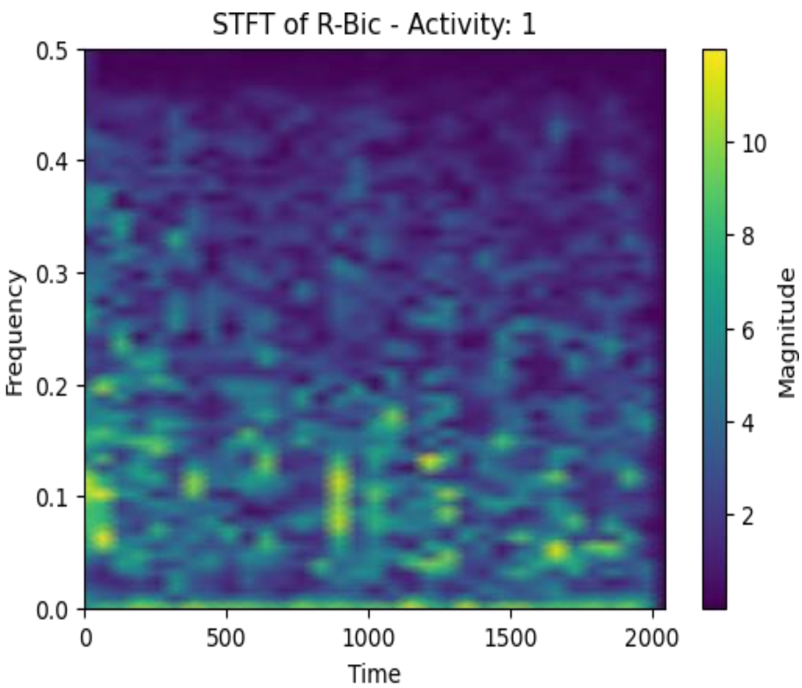


Figure 3.5: Short Time Fourier Transform of an EMG Channel

3.4 Transfer Learning

Transfer learning is a technique used in machine learning where a pre-trained model is used as a starting point for a new task. This is particularly useful when dealing with a situation that has limited data or when the new task is similar to the one the model was already trained on. Incidentally, Transfer Learning has improved our existing models fundamentally, giving better results and more in-depth analysis and significantly improving the feature extraction process. Some models are pre-trained to detect objects, so they may have an understanding and recognition ability towards edges and corners. Similarly, a huge set of pre-trained models are available which have been trained in various domains.

Transfer Learning models we used were trained on the dataset ImageNet [29]. ImageNet has a variety of images, encompassing everyday objects like furniture and clothing, the natural world with animals and plants, and even human-made structures. Each image is assigned a specific label, allowing the model to learn the visual characteristics associated with each category. ImageNet based models are being used in applications like Medical Diagnosis, Self-Driving cars and Image Search. The Transfer Learning models applied in our project are

VGG19:

A deep architecture known for its high accuracy, achieved by stacking numerous small convolutional filters. While it is powerful, VGG19 is computationally expensive due to its depth, requiring significant processing power and memory. It could be prone to overfitting with limited data; hence we need to make sure the dataset is fairly large.

MobileNet:

MobileNet is a compact convolutional neural network structure specifically created for efficient processing on mobile and embedded devices. The use of depthwise separable convolutions reduces computational complexity and memory usage while still achieving competitive accuracy. MobileNet utilizes a streamlined architecture with fewer parameters compared to traditional CNNs, making it suitable for resource-constrained environments. It offers different model sizes (e.g., MobileNetV1, MobileNetV2) with varying trade-offs between speed and accuracy, allowing flexibility for different deployment scenarios.

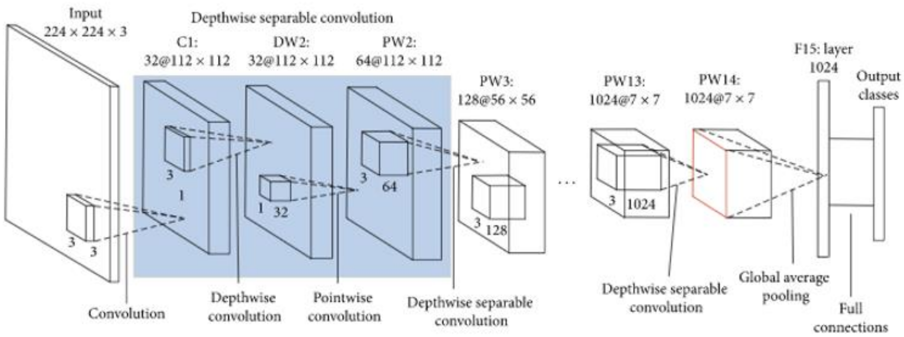


Figure 3.6: MobileNetV2 Architecture [30]

ResNet50:

The Residual Network, often known as ResNet, was developed to address the issue of the vanishing gradient problem in deep neural networks. It incorporated residual connections, which allow for the training of very deep architectures. It utilizes skip connections to add identity mappings, facilitating the flow of gradients during training and enabling the network to efficiently acquire and preserve information throughout different layers [31]. ResNet architectures are available in various depths, such as ResNet-50, ResNet-18 and ResNet-152,

DenseNet:

The Dense Convolutional Network, often known as DenseNet, incorporates dense connections between layers, allowing each layer to directly receive inputs from all preceding layers. The high level of interconnectivity between the components of the system encourages the reuse of features, improves the flow of gradients, and reduces the occurrence of the vanishing gradient problem. As a result, the learning efficiency and the propagation of features are enhanced. DenseNet architectures consist of densely connected blocks, with each block comprising convolutional, batch normalization, and activation layers.

Inception V3:

It is a part of the Inception family of architectures, employs a multi-branch structure with parallel convolutional operations of different kernel sizes and strides to capture features at multiple scales. It incorporates techniques like factorized convolutions and dimensionality reduction to reduce computational complexity while preserving representational power. Inception v3 utilizes auxiliary classifiers and batch normalization to stabilize training and enhance performance [32]. It is known for its high accuracy and efficiency in tasks such as image classification and object detection.

3.5 Synchro Squeezing transform

The SynchroSqueezing transform (SST) is a powerful signal processing technique used primarily in the analysis of non-stationary signals, where the frequency content evolves over time. It enhances the time-frequency analysis achieved through STFT by providing improved resolution and localization of spectral components. STFT breaks down a signal into short, overlapping segments, calculating the Fourier transform for each segment to analyze the signal's frequency content over time. However, traditional STFT often suffers from poor frequency resolution and difficulty in localizing frequency components accurately, especially in signals with rapidly changing frequencies.

SST addresses these limitations by re-squeezing the energy of the signal after applying STFT. It achieves this by redistributing the energy in the time-frequency domain, resulting in a more concentrated representation of spectral components. This reassignment of energy allows for better localization of instantaneous frequencies and improves the resolution of time-frequency representations [33]. By combining STFT with SynchroSqueezing, analysts can obtain a more accurate and detailed understanding of the time-varying frequency content of signals. This enhanced analysis is particularly valuable in fields such as biomedical signal processing, vibration analysis, and speech recognition, where precise characterization of frequency components over time is critical. Hence SynchroSqueezing transform with STFT (SS-STFT) offers a powerful approach for extracting meaningful information from non-stationary signals.

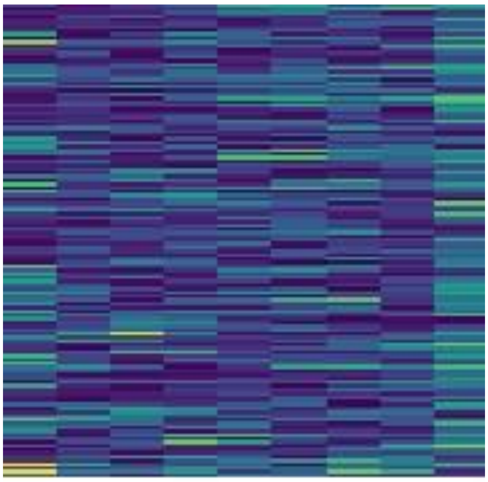


Figure 3.7: SS-STFT image for L-bicep.

Chapter 4

Methodology

The proposed methodology is as shown if Figure 4.1, it mainly consists of 3 parts, EMG physical action dataset and EMG dataset in lower limb were used as input, The aim here is to train the model to predict the activities being performed based on the EMG data provided.

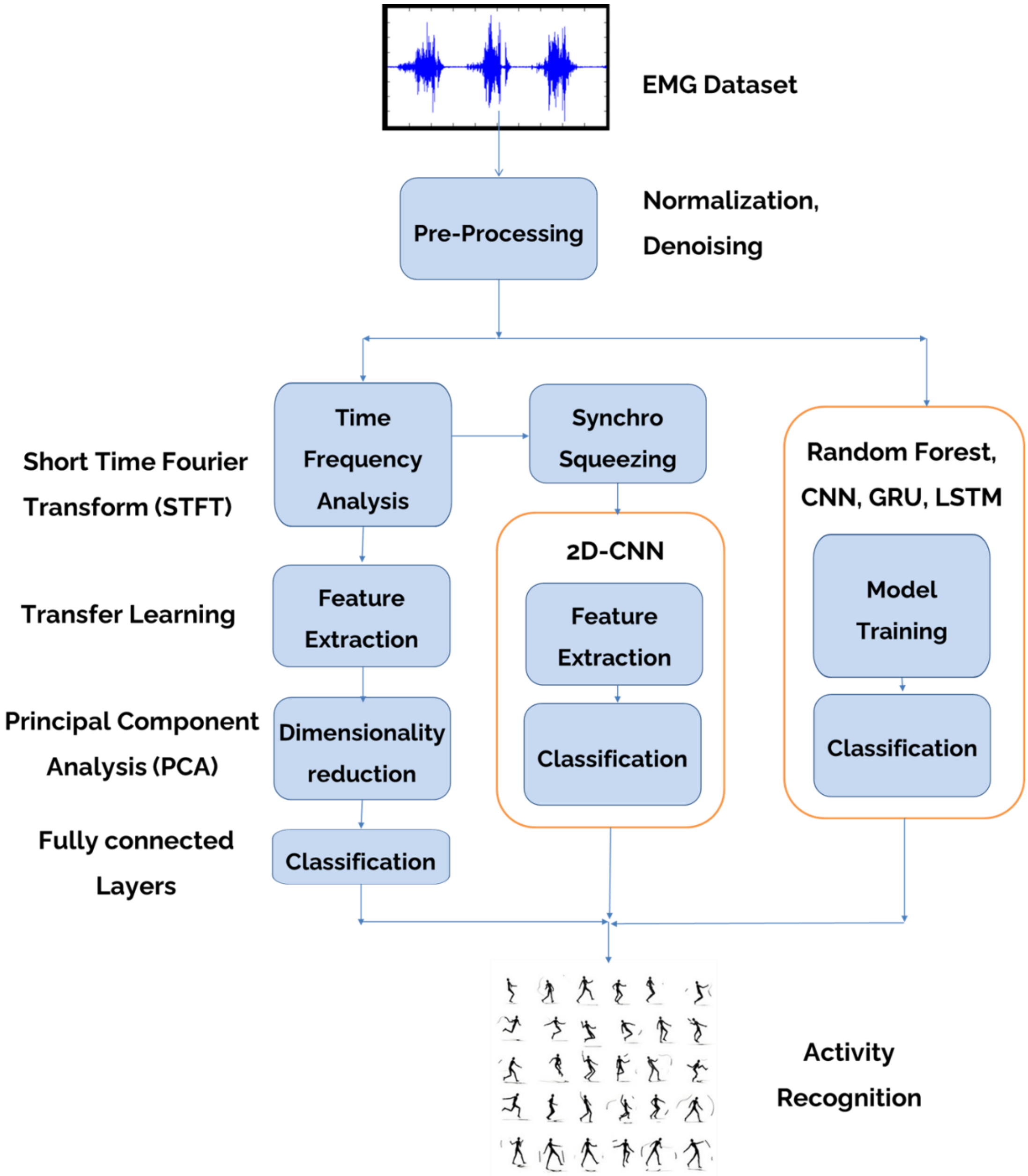


Figure 4.1: Methodology for Activity recognition using different DL models

4.1 Pre-processing

The datasets are pre-processed to organize the raw data in a much better way. An advantage of the datasets chosen is that they have no missing values or outliers. In EMG physical action dataset, there were folders named on subjects where each folder contains data in the form of txt files for the 20 activities performed, while the EMG dataset in lower limb had folders named after activities, with the txt files in it named after each subject. The following steps are followed as a part of data pre-processing:

1. **Convert text files into CSV**

- It is easy to read CSV using the "pandas" library in Python, for the conversion we used the "os" and "glob" libraries to iterate over all the subfolders and convert the txt files into CSV files. The contents of each CSV file in the datasets are as shown in the table 1, 2.

2. **Add Activity Column**

- additional column named "activity" was introduced. For the EMG physical action dataset this column assigned a value of 0 or 1 based on whether the action was classified as aggressive or normal, respectively. For the EMG dataset in lower limb values were 0 or 1 or 2 for the walking, sitting, and standing respectively.

3. **Save the CSV files subject wise**

- the CSV files were re arranged to save them as subject folders in which contained each file for the three activities, the rearrangement is done so that both the datasets are organized in the same way.

Table 1: Contents of each file in EMG physical action dataset

Column	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Muscle	R-Bic	R-Tri	L-Bic	L-Tri	R-Thi	R-Ham	L-Thi	L-Ham

Table 2: Contents of each file in EMG dataset in lower limb

Column	**1**	**2**	**3**	**4**
Muscle	Rectus Femoris (RF)	Biceps Femoris (BF)	Vastus Medialis (VM)	Semitendinosus (ST)

4.2 Random Forest and Deep Learning Models

Initially this pre-processed data is used as input to train 4 different models, they are Random Forest, CNN, LSTM and a CNN+LSTM hybrid model, The above specified models were employed to assess their effectiveness in predicting distinct physical actions based on Electromyography (EMG) signals the following procedure was followed to ensure proper training and assessment of the models:

1. **Data Compilation**

   - We consolidated all individual CSV files into a single Pandas DataFrame,

2. **Data Segmentation**

   - The dataset was bifurcated into two components: features, representing data used for predictions, and labels, indicating whether an activity is normal or aggressive. The labels can sometimes be encoded especially for deep learning models.

3. **Feature Scaling through Standardization**:

   - Features were standardized to a common scale, mitigating the dominance of any single feature during the model fitting process due to varying scales.

4. **Dataset Partitioning for Training and Testing**:

   - We partitioned the dataset into subsets for training and testing, The training subset was utilized for model instruction, while the testing subset gauged its predictive accuracy.

5. **Classifier/Model Training**:

   - Random Forest, CNN, LSTM, CNN+LSTM hybrid models were trained on the training set, the number of epochs for the deep learning models was set to 10, the learning rate was set to 0.001 to get a better accuracy but we have to make a compromise as the training time increases.

6. **Prediction, Accuracy and Confusion Matrix**:

- Models predicted activities on the testing subset, confusion matrix summarizes the predictions of a model on a dataset by comparing the actual labels with the predicted labels The analysis offers valuable information regarding the model's accuracy, precision, recall, and F1-score for various classes. helping to identify where the model is making errors, such as misclassifications or biases towards certain classes.

4.3 Transfer Learning Models

Spectrograms are created using the Short Time Fourier Transform (STFT), which utilizes the same pre-processed data, these spectrograms are given as input for five different transfer learning models, VGG19, MobileNet, DenseNet, ResNet50, InceptionV3, To implements the steps below are followed and are also depicted in Figure 4.3.

1. **Load the transfer learning model:**

- from the TensorFlow library in python ImageNet transfer learning model (weights) is imported without the top classification layers.

2. **Feature Extraction:**

- The time frequency images generated with the help of STFT are given as input, these images are stored class wise, we iterate over each image load it, pre-process it to match the input requirements of the transfer learning model that is used.

- The transfer learning model being pre-trained extracts high level features form the input image. These features capture various patterns, textures and structures present in the image. The extracted features are matched along with their corresponding class labels.

- After extracting features from all images, the extracted features are concatenated into a single array. Meanwhile, the class labels are converted to numeric format for easier handling in machine learning algorithms. This conversion is done using a dictionary where each unique label is mapped to a numeric value.

3. **Flatten the extracted features:**

- The extracted features are typically in the form of a multi-dimensional array where each image has a set of features. To perform dimensionality reduction or use these features in machine learning models, it's necessary to flatten these features into a 2D array where each row represents an image and each column represents a feature.

4. **Principal component analysis (PCA):**

- Principal Component Analysis (PCA) is an unsupervised learning approach employed in machine learning to reduce the dimensionality of data. Principal Component Analysis (PCA) is a statistical technique that transforms correlated observations into a set of linearly uncorrelated characteristics using an orthogonal transformation.

- The Principal Components refer to the newly changed features. This is especially advantageous when working with complex data sets that have a large number of dimensions, making analysis challenging.

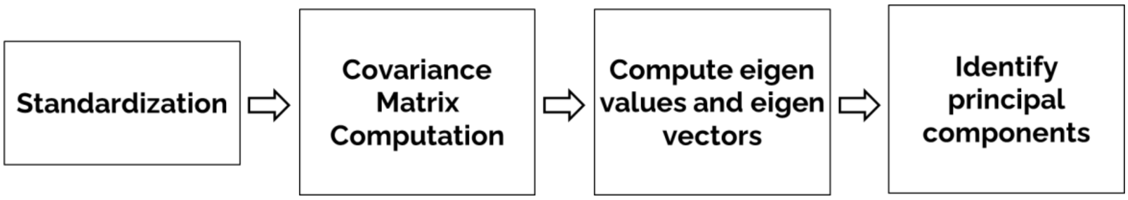


Figure 4.2: Steps to perform PCA

5. **Fully connected layers:**

- Dense layers, also referred to as fully connected layers, establish connections between every neuron in one layer and every neuron in the subsequent layer, allowing for intricate feature extraction. Within these layers, every neuron performs a calculation by multiplying the inputs from the preceding layer by specific weights, and then applies an activation function.

- They play a vital role in learning patterns and relationships in the input data, making them essential for tasks like classification and regression. Fully connected layers are a prevalent component in different types of neural network topologies, such as feedforward, convolutional, and recurrent neural networks. Their flexibility and ability to learn intricate representations contribute significantly to the success of deep learning models.

- The reduced features after PCA are first split into testing and training sets and fully connected layers of 3 different models were used (CNN, GRU, LSTM) for classification, the model is trained with the training data set and testing dataset is used to predict the classes, for each of the transfer learning models.

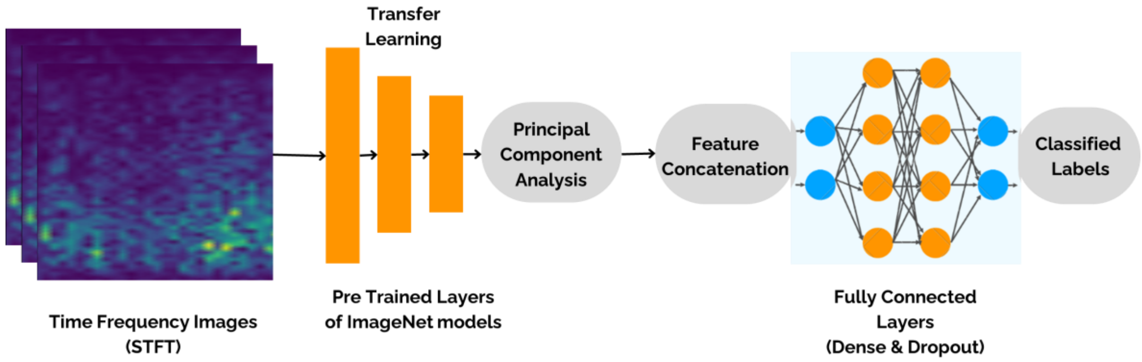


Figure 4.3: Steps to implement a transfer learning-based model

4.4 2D-CNN with STFT and SS-STFT images as input

The original time series EMG data is transformed into time frequency images with the help of STFT function in SciPy library of python, on these STFT images SynchroSqueezing transform was applied and both the time frequency images are used to train a specific 2D-CNN model whose model summary is as shown in figure 4.4, CNN is used because it is best suited for tasks such as image classification, object detection, and image segmentation, thanks to their ability to automatically learn and extract meaningful features from visual data. The images generated are stored class wise, in case of EMG physical action dataset it is two folders one for normal actions, the other for aggressive. The steps after that are test- train split, model training, prediction and accuracy assessment, the learning rate considered is 0.001, and the number of epochs run when STFT based time frequency images are give as input are 10, while when the input is SS-STFT based time frequency images it is 50.

Layers(type)	**Output Shape**	**Param#**
conv2d (Conv2D)	(None,222,222,32)	896
Max_pooling2d (MaxPooling2D)	(None,111,111,32)	0
conv2d_1(Conv2D)	(None,109,109,64)	18 496
Max_pooling2d_1(MaxPooling2D)	(None,54,54,64)	0
conv2d_2Conv2D)	(None,52,52,128)	73 856
flatten(Flatten)	(None,346112)	0
dense (Dense)	(None,128)	44 302 464
dropout(Dropout)	(None,128)	0
dense_1(Dense)	(None,2)	258
Total params: 44 395 970 **Trainable params: 44 395 970** **Nontrainable params: 0**		

Figure 4.4: Model summary of 2D-CNN

Chapter 5

Results and Discussion

All the deep learning models mentioned in the methodology were implemented on a workstation running on 64-bit windows 10 operating system, with the following specs:

Processor details: Intel Xeon W – 2133 @ 3.60 GHz [34]

RAM: 64 GB DDR4

GPU: NVIDIA Quadro P4000 [35]

GPU SIZE: 8 GB, DDR5

Anaconda Navigator was used as the graphical user interface as it is easy to manage environments and is an open-source platform for Python, Jupyter Notebook version 6.5.2 was used as the Integrated development environment (IDE). Codes were run as cells in a IPython Notebook (IPYNB) which is a document format used for interactive computing in Python

For time frequency images generated for both STFT and SS-STFT were 9600 in total and learning rate considered is 0.001, and the number of epochs run when STFT based time frequency images are given as input to CNN are 10, while when the input is SS-STFT based time frequency images are given as in put to CNN is 50. For MobileNet implemented on EMG physical action dataset and EMG dataset in lower limb a minimum of 20 epochs were run to classify the activities with the help of final fully connected layers. For any other model the default number of epochs run were set to 10.

5.1 Evaluation Metrics

To evaluate different deep learning and transfer learning models, we use various metrics such as accuracy, precision, recall, F1-score, to understand how to derive these metrics first we need to know what are Ture Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), for a binary classification models such as activity classification on EMG physical action dataset, the activity can be either aggressive or normal depending on the actual and predicted labels of the activity TP,TN, FP, FN can be defined as shown in Table 3.

Table 3: Definition of TP, TN, FP, FN

Predicted Label	Actual Label	Definition
Positive	Positive	**True Positive (TP)**
Positive	Negative	**False Positive (FP)**
Negative	Positive	**False Negative (FN)**
Negative	Negative	**True Negative (TN)**

1. **Accuracy:** Accuracy is a frequently used statistic in ordinary conversation. It represents the proportion of properly anticipated outcomes out of the total number of outcomes. Precision is a rudimentary metric and can occasionally be deceptive.

$$\text{Accuracy} = \frac{TP+TN}{TP + FP+TN+FN}$$

2. **Precision:** Precision calculates the ratio of correctly identified positive instances to the total number of positive instances predicted by the model.

$$\text{Precision} = \frac{TP}{TP + FP}$$

3. **Recall:** Recall or sensitivity, refers to the proportion of correctly identified positive instances.

$$\text{Recall} = \frac{TP}{TP + FN}$$

4. **F1-Score:** It can be utilized to gauge the efficiency of our models in balancing precision and recall. The F1 score has a crucial characteristic: when either precision or recall or both are zero then it is zero it will be zero

$$\text{F1-Score} = \frac{2 \cdot Precision.Recall}{Precision+Recall}$$

5.2 Random Forest & Deep learning models

Table 4 below provides the values of TP, FP, FN, TN followed by accuracy, precision, recall, F1-score of Random Forest (RF), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), a hybrid of model of CNN+LSTM.

Table 4: Accuracy and other metrics of deep learning models

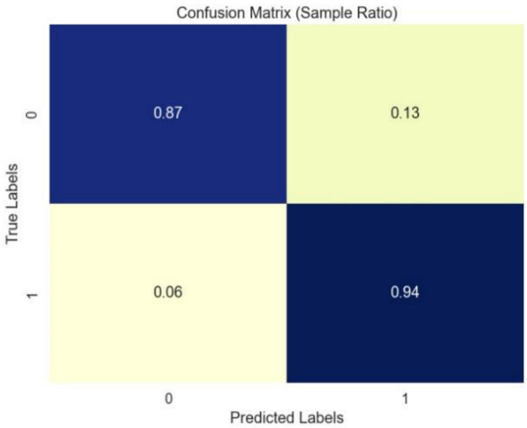
Model	TP	FP	FN	TN	Accuracy	Precision	Recall	F1 SCORE
RF	0.87	0.13	0.06	0.94	0.905	0.87	0.9354	0.9015
CNN	0.82	0.18	0.15	0.85	0.835	0.82	0.8454	0.8325
LSTM	0.87	0.13	0.10	0.90	0.885	0.87	0.8969	0.8832
CNN+LSTM	**0.89**	**0.11**	**0.07**	**0.93**	**0.91**	**0.89**	**0.9270**	**0.9081**

Initially, a simple Machine learning model i.e., Random Forest was implemented on the data which could effectively handle diverse EMG signal features resulting in an accuracy of 90%. It also delivered a Precision of 87% and a recall rate of 93%, combined, the F1 score which evaluates both the precision and recall is 0.90. The CNN model gave an accuracy of 83% with a precision of 82% and a recall value of 84%, combined resulting in an F1 score of 0.83.

The LSTM model gave an accuracy of 88% with a precision of 87% and a recall value of 89%, resulting in an F1 score of 0.88, which is comparatively higher than the CNN mode. The reason being that LSTM models are much better with sequential data.

The hybrid CNN+LSTM model produced a higher accuracy than both of them individually, 91% with a precision of 89% and a recall value of 92%. The F1 score 0.90 which is also higher than the individual models, indicating higher precision and recall balance.

It can be inferred that CNN+LSTM model is much better suited on the data than random forest or CNN, LSTM alone. The reason for using CNN+LSTM as a hybrid model is that it leverages the advantages of both CNN and LSTM networks in processing sequential input.



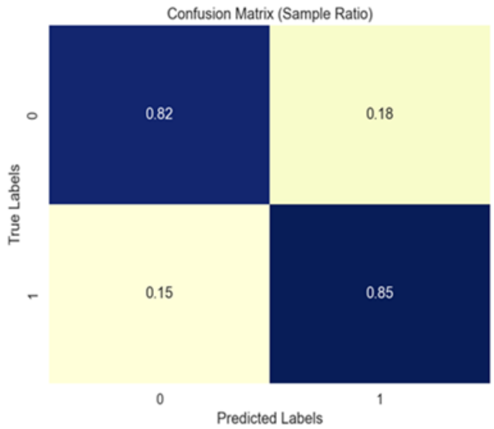
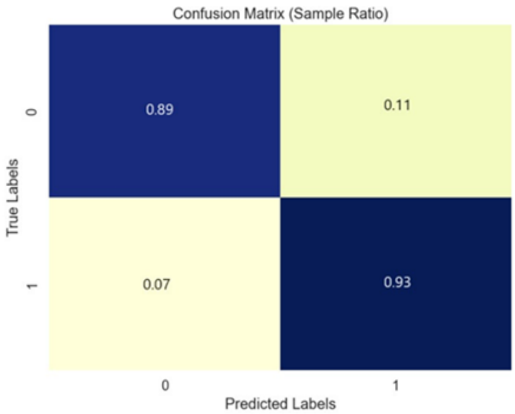


Figure 5.1, 5.2: Confusion Matrices of Random Forest and CNN respectively



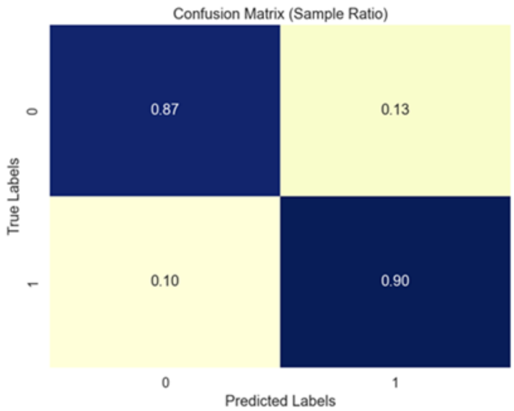


Figure 5.3, 5.4: Confusion Matrices of CNN+LSTM and LSTM respectively

5.3 Transfer Learning Models

Giving the STFT based time frequency images as input to various transfer learning models we have achieved much higher accuracies than the CNN model discussed at the end of previous section. In total Five Transfer learning models along with Three different fully connected layers have been tested and their results have been compiled in Table 5.

Table 5: Accuracy and other metrics of various TL Models and Fully Connected layers

Model	Number of features extracted	Time Taken for each step (ms)	Fully connected Layers Accuracy		
			CNN	LSTM	GRU
MobileNet	**50176**	**50**	**97.34**	**96.71**	**96.82**
VGG19	25088	150	95.67	95.88	95.68
DenseNet	50716	150	97.13	96.45	96.14
ResNet50	100352	150	97.03	97.08	96.45
InceptionV3	131072	170	96.71	96.04	96.14

- Data trained on **MobileNet** and fed through the CNN Layer gave an accuracy of 97.3% and it extracted 50,176 features in total, taking 50 ms/step. Similarly, LSTM gave an accuracy of 96.71% and an accuracy of 96.82% with GRU.

- Data trained on **VGG19** and fed through the CNN Layer gave an accuracy of 95.67% and it extracted 25,088 features in total, taking 150 ms/step, i.e., half the features as MobileNet and triple the time per step taken in MobileNet. Similarly, LSTM gave an accuracy of 95.88% and GRU gave an accuracy of 95.68%.

- Data trained on **DenseNet** and fed through the CNN Layer gave an accuracy of 97.13% and it extracted 50,716 features in total, taking 150 ms/step. Similarly, LSTM gave an accuracy of 96.45% and an accuracy of 96.14% with GRU. Which means the accuracies were more or less the same as MobileNet but the time taken per step was triple that of MobileNet.

- Data trained on **ResNet50** and fed through the CNN Layer gave an accuracy of 97.03% and it extracted 1,00,352 features in total taking 150 ms/step. Similarly, LSTM gave an accuracy of 97.08% and an accuracy of 96.45% with GRU. Clearly, the accuracy and the time per step are similar to DenseNet and VGG but the Features extracted are almost double that of MobileNet.
- Data trained on **InceptionV3** and fed through the CNN Layer gave an accuracy of 96.71% and it extracted 1,31,072 features in total taking 170 ms/step. Similarly, LSTM gave an accuracy of 96.04% and an accuracy of 96.14% with GRU. It can be concluded that InceptionV3 extracted the highest number of features, but also consumed the highest amount of time per step while maintaining the accuracies around the same range.

Nevertheless, different models have worked best with different fully connected layers. The highest accuracy among all models with the CNN layer is from MobileNet, whereas that with LSTM is given by Resnet50, and with GRU we see that the highest accuracy was achieved by MobileNet. In terms of the highest number of features extracted, InceptionV3 leads by extracting 1,31,072 features. In terms of lowest time taken per step, MobileNet takes the lead with 50 ms/step. **It is clear from the observations and discussions that MobileNet gives the highest accuracies with a decent number of features extracted and with the most minimal amount of time per step.**

5.4 EMG Dataset in lower limb

To verify the results and ensure the MobileNet model is not overfitting we tested it for another dataset, The EMG Dataset in lower limb. The results are as displayed in Table 6.

Table 6: Accuracy of various fully connected layers with different datasets

Dataset	Fully Connected Layers (MobileNet)		
	CNN	**LSTM**	**GRU**
EMG Physical action dataset	98.85	98.85	97.91
EMG dataset in lower limb	92.42	92.80	93.56

The EMG Physical action dataset delivered an accuracy of 98.85% with a CNN model, 98.85% with an LST model and a 97.91% with a GRU model, while the EMG dataset in lower limb gave an accuracy of 92.42% with CNN, 92.80% with LSTM and 93.56% with GRU. **This proves the fact that the MobileNet model works good for other EMG datasets as well providing satisfactory accuracy. The decrease in accuracy is due to the increase in number of classes in EMG dataset in lower limb compared to the EMG Physical action dataset.**

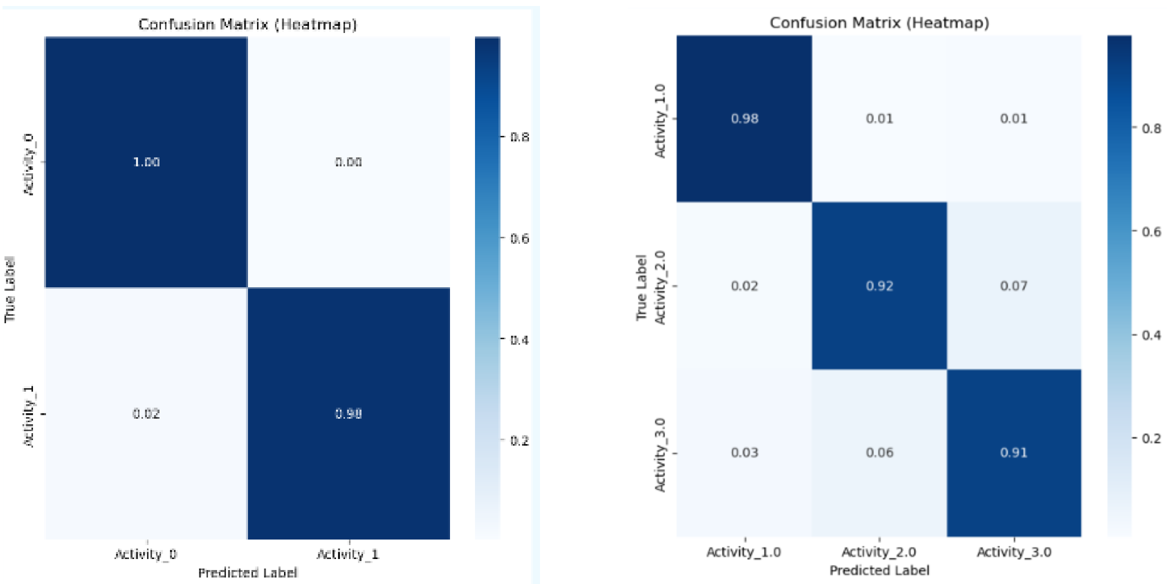


Figure 5.5, 5.6: Confusion Matrices for EMG Physical Action & EMG dataset in lower limb respectively

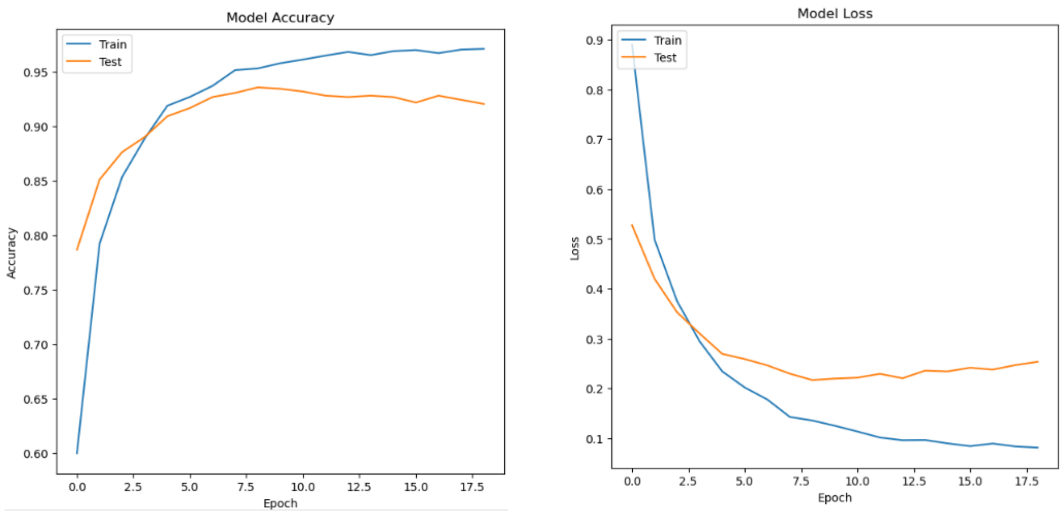


Figure 5.7, 5.8: Training and Testing accuracy per epoch & Loss per epoch graphs respectively for EMG dataset in lower limb

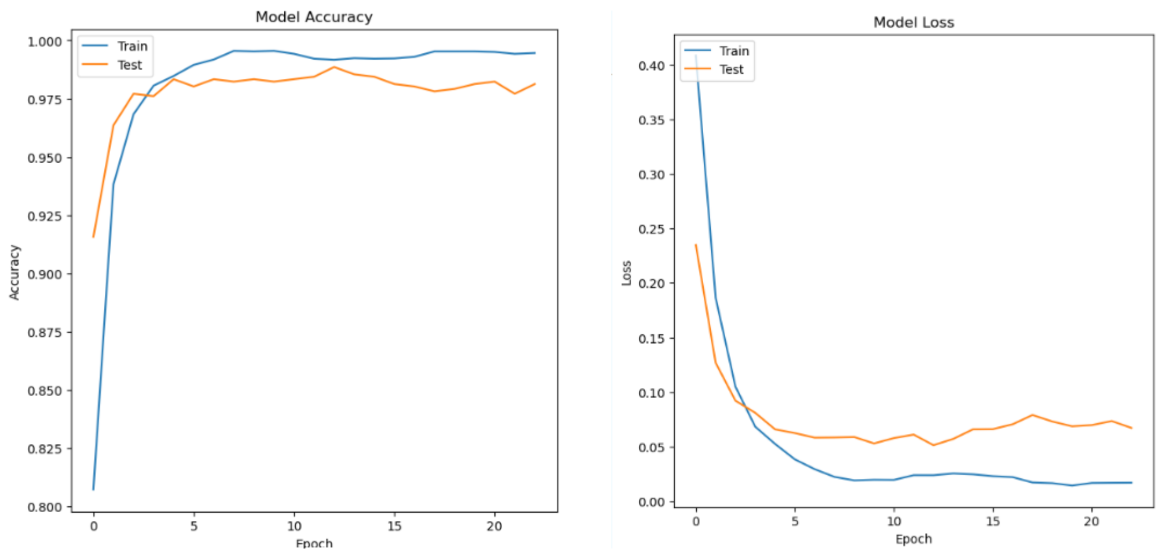


Figure 5.9, 5.10 Training and Testing accuracy per epoch & Loss per epoch graphs respectively for the EMG Physical Action Dataset

5.5 STFT and Synchro Squeezing transform (SST)

Table 7 below provides the evaluation metrics of 2D-CNN model with input as images of STFT and SS-STFT.

Table 7: Accuracy and other metrics of 2D-CNN model

Input	TP	FP	FN	TN	Accuracy	Precision	Recall	F1 SCORE
STFT	0.96	0.04	0.06	0.94	0.95	0.96	0.9411	0.9504
SS-STFT	**0.98**	**0.02**	**0.01**	**0.99**	**0.985**	**0.98**	**0.9898**	**0.9848**

STFT was applied to derive time frequency images of the EMG data. This greatly aided the feature extraction process of the model. These images were given as input to a 2D-CNN to give an accuracy of 95% with precision and recall of 96% and 94%, resulting in a F1 score of 0.95, following the STFT SynchroSqueezing was also applied on the images, the new images generated were used as input for 2D-CNN, it gave a staggering accuracy of 98.5% almost as good as the transfer learning models with STFT as input, with precision and recall of 98% and 98.98%, resulting in a F1 score of 0.9848.This shows that SS-STFT is much better than STFT in case of generating time frequency images for activity recognition.

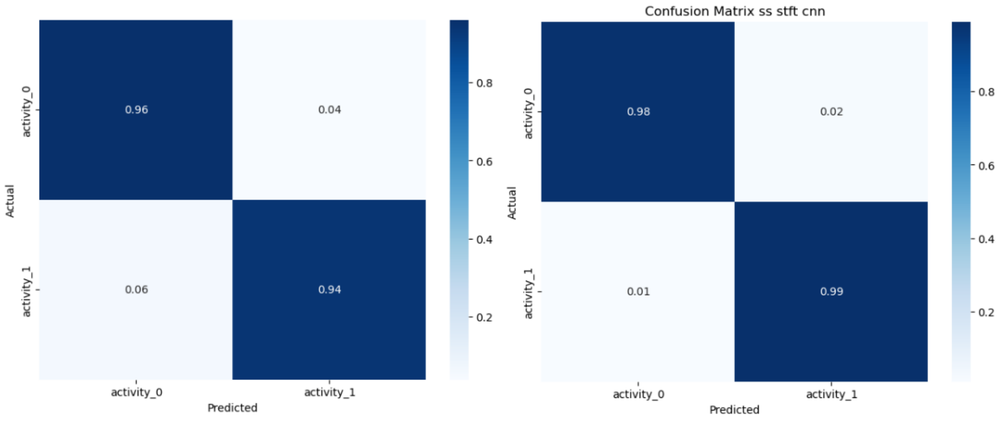
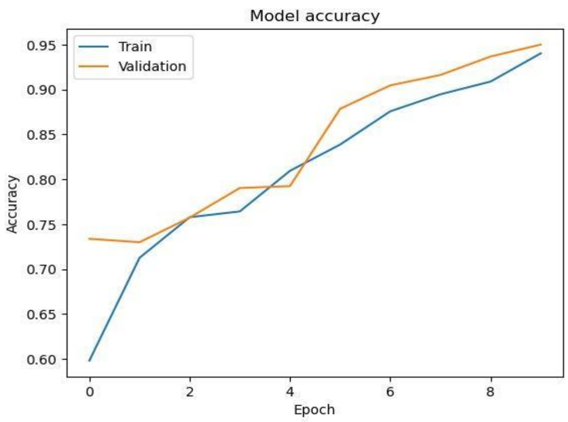


Figure 5.11&5.12: Confusion matrices of 2D-CNN with STFT & SS-STFT as input respectively



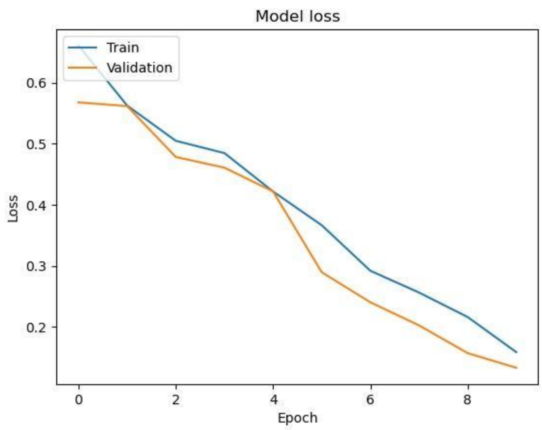
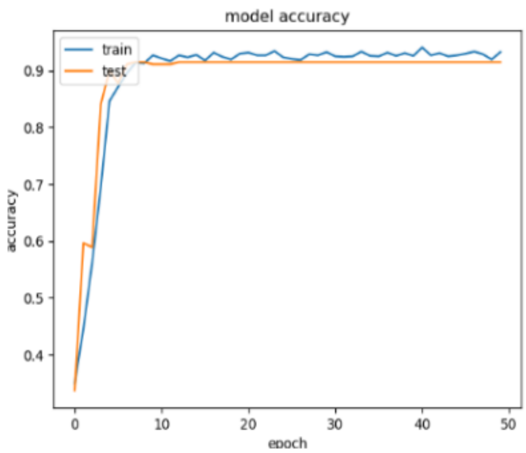


Figure 5.13 & 5.14: Training and Testing accuracy per epoch & Loss per epoch graphs respectively for 2D-CNN with STFT images as input



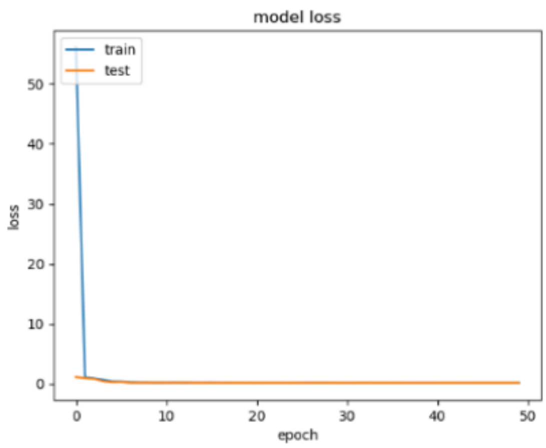


Figure 5.15 & 5.16: Training and Testing accuracy per epoch & Loss per epoch graphs respectively for 2D-CNN with SS-STFT images as input

Chapter 6

Conclusion and Future Scope

Based on the results obtained and comparing different deep learning and transfer learning model trained with the help of datasets from the University of California, Irvine (UCI) we can conclude that for raw EMG data in sequential time series format CNN+LSTM model worked the best according to traditional Machine Learning approaches, with an accuracy of 91%. It combined the best parts of both the models, later this sequential data was used to create time frequency images with the help of Short-Time Fourier Transform (STFT). When it comes to image classification CNN's gave the best results when STFT images were given as input, achieving an accuracy of 95%. It was derived that CNNs excel at analyzing spatial relationships within images. In an STFT image, these spatial features correspond to the distribution of frequencies across time.

Moving further we employed Transfer Learning models on both the EMG physical action dataset and EMG dataset in lower limb. MobileNet provided the best results with efficient, light weight and high accuracies on both the datasets without overfitting. The MobileNet + CNN model gave an accuracy of 97.34%. The other models used were VGG19, ResNet50, DenseNet, Inception V3. Mobilenet was designed as a lightweight architecture, trained on a huge ImageNet dataset which provided a balance on computational efficiency.

The combination of STFT+SS+CNN produced the most promising results for lower limb muscle activity analysis through sEMG signals, achieving an impressive 98.56% accuracy on the dataset. The combination of STFT+SS+CNN is robust to temporal dynamics, ensuring accurate predictions. The model interpretability allows for insights into learned features, enhancing its trustworthiness. This hybrid approach demonstrated superior capabilities in feature extraction and classification. By leveraging the strengths of STFT-SS for practical time–frequency analysis and CNN for spatial feature extraction, this model provides a robust foundation for future predictions and diagnoses.

Further work in this project can be done in the following areas:

- **Real time data collection**

  - collection of real time EMG data form subjects with Sarcopenia from hospital to increase the prediction accuracy of the developed model. This would allow for more dynamic and continuous, real time activity recognition through a well trained model.

- **Denoising of EMG signal**

  - sEMG signals can be affected by a variety of noise sources, like interference from other bioelectrical signals (EEG, ECG) or equipment interference. To improve the model's effectiveness, we can implement sophisticated denoising techniques. This could involve methods like wavelet transform denoising, adaptive filtering, or deep learning-based approaches to remove noise while preserving the integrity of the sEMG signal.

- **Testing of computationally intensive transfer learning models**

  - The current project utilizes transfer learning models with relatively low computational demands. However, exploring models with higher computational intensity could potentially lead to further performance gains.

  - Evaluating the performance of computationally intensive models like AlexNet and GoogleNet. These models have demonstrated high accuracy in image recognition tasks, and their transfer learning capabilities could be harnessed for sEMG-based activity prediction. While these models require more processing power, advancements in hardware and cloud computing might make them increasingly feasible for practical applications.

References

[1] Theodoridis,Theo. (2011). EMG Physical Action Data Set. UCI Machine Learning Repository. https://doi.org/10.24432/C53W49.

[2] Sanchez,Oscar and Sotelo,Jose. (2014). EMG dataset in Lower Limb. UCI Machine Learning Repository. https://doi.org/10.24432/C5ZW3P.

[3] Kumar, P., Suresh, S. Deep-HAR: an ensemble deep learning model for recognizing the simple, complex, and heterogeneous human activities. Multimed Tools Appl 82, 30435–30462 (2023). https://doi.org/10.1007/s11042-023-14492-0

[4] Ferrari, A., Micucci, D., Mobilio, M., Napoletano, P. (2019). Human Activities Recognition Using Accelerometer and Gyroscope. In: Chatzigiannakis, I., De Ruyter, B., Mavrommati, I. (eds) Ambient Intelligence. AmI 2019. Lecture Notes in Computer Science(), vol 11912. Springer, Cham. https://doi.org/10.1007/978-3-030-34255-5_28

[5] Zhang, Shibo, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and Nabil Alshurafa. 2022. "Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances" Sensors 22, no. 4: 1476. https://doi.org/10.3390/s22041476

[6] F. Auger et al., "Time-Frequency Reassignment and Synchrosqueezing: An Overview," in IEEE Signal Processing Magazine, vol. 30, no. 6, pp. 32-41, Nov. 2013, doi: 10.1109/MSP.2013.2265316.

[7] S. Wang and B. Chen, "Split-Stack 2D-CNN for Hand Gestures Recognition Based on Surface EMG Decoding," 2020 Chinese Automation Congress (CAC), Shanghai, China, 2020, pp. 7084-7088, doi: 10.1109/CAC51589.2020.9327333.

[8] Inhyuk Moon, Myungjoon Lee, Junuk Chu and Museong Mun, "Wearable EMG-based HCI for Electric-Powered Wheelchair Users with Motor Disabilities," Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 2005, pp. 2649-2654, doi: 10.1109/ROBOT.2005.1570513.

[9] Alaskar H. Deep learning of EMG time-frequency representations for identifying normal and aggressive actions. IJCSNS Int. J. Comput. Sci. Netw. Secur. 2018 Dec 30;18(10.1142):S2424922X18400065.

[10] A. Vijayvargiya, B. Singh, R. Kumar, U. Desai, and J. Hemanth, ‘Hybrid Deep Learning Approaches for sEMG Signal-Based Lower Limb Activity Recognition’, Mathematical Problems in Engineering, vol. 2022, p. 3321810, Nov. 2022.

[11] Demir, Fatih & Bajaj, Varun & Ince, Melih & Taran, Sachin & Sengur, Abdulkadir. (2019). Surface EMG signals and deep transfer learning-based physical action classification. Neural Computing and Applications. 31. 10.1007/s00521-019-04553-7.

[12] Costa-Garcia A, Murai A, Shimoda S. Pattern Matching for Real-Time Extraction of Fast and Slow Spectral Components From sEMG Signals. IEEE Trans Neural Syst Rehabil Eng. 2023;31:3587-3596. doi: 10.1109/TNSRE.2023.3311037. Epub 2023 Sep 12. PMID: 37656649.

[13] Hassan MM, Huda S, Uddin MZ, Almogren A, Alrubaian M. Human Activity Recognition from Body Sensor Data using Deep Learning. J Med Syst. 2018 Apr 16;42(6):99. doi: 10.1007/s10916-018-0948-z. PMID: 29663090.

[14] Mekruksavanich, Sakorn & Jantawong, Ponnipa & Jitpattanakul, Anuchit. (2022). A Deep Learning-based Model for Human Activity Recognition using Biosensors embedded into a Smart Knee Bandage. Procedia Computer Science. 214. 621-627. 10.1016/j.procs.2022.11.220.

[15] Fard Moshiri, Parisa, Reza Shahbazian, Mohammad Nabati, and Seyed Ali Ghorashi. 2021. "A CSI-Based Human Activity Recognition Using Deep Learning" Sensors 21, no. 21: 7225. https://doi.org/10.3390/s21217225

[16] Ronao, Charissa & Cho, Sung-Bae. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications. 59. 10.1016/j.eswa.2016.04.032.

[17] Khan, Md & Moznuzzaman, Md & Ide, Shuya. (2023). Analysis of aging effect on lower limb muscle activity using short time Fourier transform and wavelet decomposition of electromyography signal. AIP Advances. 13. 10.1063/5.0148044.

[18] Tao, Wenjin & Lai, Ze-Hao & Leu, Ming & Yin, Zhaozheng. (2018). Worker Activity Recognition in Smart Manufacturing Using IMU and sEMG Signals with Convolutional Neural Networks. 26. 1159-1166. 10.1016/j.promfg.2018.07.152.

[19] Li, Chuanjiang & Deng, Huiyin & Yin, Shiyi & Wang, Chenming & Zhu, Yanfei. (2023). sEMG signal filtering study using synchrosqueezing wavelet transform with differential evolution optimized threshold. Results in Engineering. 18. 101150. 10.1016/j.rineng.2023.101150.

[20] Xi, Xugang, Minyan Tang, Seyed M. Miran, and Zhizeng Luo. 2017. "Evaluation of Feature Extraction and Recognition for Activity Monitoring and Fall Detection Based on Wearable sEMG Sensors" Sensors 17, no. 6: 1229. https://doi.org/10.3390/s17061229

[21] Ding, Zhen, Chifu Yang, Zhihong Tian, Chunzhi Yi, Yunsheng Fu, and Feng Jiang. 2018. "sEMG-Based Gesture Recognition with Convolution Neural Networks" Sustainability 10, no. 6: 1865. https://doi.org/10.3390/su10061865

[22] Bansal, Hunish & Chinagundi, Basavraj & Rana, Prashant & Kumar, Neeraj. (2022). An Ensemble Machine Learning Technique for Detection of Abnormalities in Knee Movement Sustainability. Sustainability. 14. 13464. 10.3390/su142013464.

[23] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

[24] https://www.javatpoint.com/machine-learning-random-forest-algorithm

[25] Phung, & Rhee,. (2019). A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. Applied Sciences. 9. 4500. 10.3390/app9214500.

[26] Yong Yu, Xiaosheng Si, Changhua Hu, Jianxun Zhang; A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Comput 2019; 31 (7): 1235–1270. doi: https://doi.org/10.1162/neco_a_01199

[27] https://www.simplilearn.com/tutorials/artificial-intelligence- tutorial/lstm#:~:text=LSTMs%20are%20able%20to%20process,problem%20that%20plagues%20traditional%20RNNs.

[28] T. Bao, S. A. R. Zaidi, S. Xie, P. Yang and Z. -Q. Zhang, "A CNN-LSTM Hybrid Model for Wrist Kinematics Estimation Using Surface Electromyography," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-9, 2021, Art no. 2503809, doi: 10.1109/TIM.2020.3036654.

[29] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848

[30] https://www.pluralsight.com/resources/blog/guides/transfer-learning-in-deep-learning-using-tensorflow-20

[31] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON), Bengaluru, India, 2021, pp. 96-99, doi: 10.1109/CENTCON52345.2021.9687944

[32] Wang, Xiangrui, Lu Tang, Qibin Zheng, Xilin Yang, and Zhiyuan Lu. 2023. "IRDC-Net: An Inception Network with a Residual Module and Dilated Convolution for Sign Language Recognition Based on Surface Electromyography" Sensors 23, no. 13: 5775. https://doi.org/10.3390/s23135775

[33] Lin Li, Haiyan Cai, Hongxia Han, Qingtang Jiang, Hongbing Ji,Adaptive short-time Fourier transform and synchrosqueezing transform for non-stationary signal separation, Signal Processing, Volume 166, 2020, 107231, ISSN 0165-1684, https://doi.org/10.1016/j.sigpro.2019.07.024.

[34] https://www.intel.com/content/www/us/en/products/sku/125040/intel-xeon-w2133-processor-8-25m-cache-3-60-ghz/specifications.html

[35] https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/products page/quadro/quadro-desktop/quadro-pascal-p4000-data-sheet-a4-nvidia-704358-r2-web.pdf