ELDERLY INDIVIDUAL FALL DETECTION SYSTEM USING DEEP LEARNING

Thesis Submitted
In Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
Software Engineering

Submitted by

Piyush Agarwal
(2K23/SWE/10)

Under the Supervision of
Dr. Abhilasha Sharma
(Associate Professor, SE, DTU)

Department of Software Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

May 2025

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to **Dr. Abhilasha Sharma**, Assistant Professor at the Department of Software Engineering, Delhi Technological University, for her invaluable guidance and unwavering encouragement throughout this research. Her vast knowledge, motivation, expertise, and insightful feedback have been instrumental in every aspect of preparing this research plan.

I am also grateful to **Prof. Ruchika Malhotra**, Head of the Department, for her valuable insights, suggestions, and meticulous evaluation of my research work. Her expertise and scholarly guidance have significantly enhanced the quality of this thesis.

My heartfelt thanks go out to the esteemed faculty members of the Department of Software Engineering at Delhi Technological University. I extend my gratitude to my colleagues and friends for their unwavering support and encouragement during this challenging journey. Their intellectual exchanges, constructive critiques, and camaraderie have enriched my research experience and made it truly fulfilling.

While it is impossible to name everyone individually, I want to acknowledge the collective efforts and contributions of all those who have been part of this journey. Their constant love, encouragement, and support have been indispensable in completing this M.Tech thesis.

Piyush Agarwal

(2K23/SWE/10)

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE DECLARATION

I PIYUSH AGARWAL hereby certify that the work which is being presented in the thesis entitled **Elderly Individual Fall Detection System using Deep Learning** in partial fulfillment of the requirements for the award of the Degree of Master of Technology submitted in the Department of Software Engineering, Delhi Technological University in an authentic record of my work carried out during the period from August 2023 to May 2025 under the supervision of Dr. Abhilasha Sharma.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Piyush Agarwal

This is to certify that the student has incorporated all the corrections suggested by the examiner in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor(s) **Signature of External Examiner**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

I hereby certify that **Piyush Agarwal (Roll no 2K23/SWE/10)** has carried out their research work presented in this thesis entitled **"Elderly Individual Fall Detection System using Deep Learning"** for the award of **Master of Technology** from the Department of Software Engineering, Delhi Technological University, Delhi under my supervision. The thesis embodies results of original work, and studies are carried out by the student herself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

Dr. Abhilasha Sharma

Associate Professor

Department of Software Engineering

DTU-Delhi, India

Elderly Individual Fall Detection System using Deep Learning

Piyush Agarwal

ABSTRACT

Falls often harm the elderly, end in serious injury, may require hospital treatment and unfortunately can be deadly. Ensuring elderly people receive rapid medical care following a slip or fall has become more difficult now that there are people that are more elderly. For this reason, there is now a greater desire for systems that can automatically spot falls to give swift care. In this research work, a CNN-BiLSTM based approach using skeletal key points for elderly fall detection is proposed. Our process includes two parts: first, we use Google's MoveNet Thunder to extract the body pose from each video frame and then, we pass the series of skeletal keypoints through a network consisting of Convolutional Neural Networks for understanding spatial features and Bidirectional Long Short Term Memory networks for learning the motion itself. The two-stage process makes it possible for the model to recognize falls accurately by observing both how someone moves and his or her body position in the sequence. Publicly available benchmark dataset, URFD, was utilized for experiments in this work. No standard guidelines have been made for the design of falling detection systems, despite their many forms. For this reason, we have gathered a range of works for this topic to provide a summary of where research on human position based fall detection algorithms currently is. Based on the URFD dataset, the proposed hybrid approach performs better than existing methods, mainly because it can represent complex fall events well. We have found that the model obtains an accuracy of 95.80%, a precision of 93.90%, a recall of 94.78%, a specificity of 93.75% and an AUC of 0.9311.

Keywords: CNN, BiLSTM, Deep Learning, Fall Detection, Elderly Care

TABLE OF CONTENTS

LIST OF TABLE(S)

LIST OF FIGURE(S)

LIST OF ABBREVIATION(S)

AI	Artificial Intelligence
BiLSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
DL	Deep Learning
FDD	Fall Detection Dataset
FPD	Fast Pose Detection
FPN	Feature Pyramid Network
HAR	Human Activity Recognition
IOT	Internet of Things
KNN	K Nearest Neighbor
ML	Machine Learning
MLP	Multilayer Perceptron
NLP	Natural Language Processing
NMS	Non Maximum Suppression
NN	Neural Network
RF	Random Forest
SVM	Support Vector Machine
URFD	University of Rzeszow Fall Dataset
VGG	Visual Geometry Group
YOLO	You Only Look Once

CHAPTER 1

INTRODUCTION

This research work addresses the usages of various Deep Learning (DL) models in the Fall Detection. In addition, our research emphasize the value of multi-model fusion and imply that integrating several learning modalities can have positive effects in the computer vision.

1.1 Background

A recent United Nations [1] report points out that there is an accelerating increase of the aged population across the globe. According to estimates, the number of older people will reach 1.5 billion by 2050, up from 727 million in 2020. This demographic transition will increase the proportion of older persons from 9.3% in the year 2020 to 16% in 2050. The older population is growing fast due to increasing longevity caused by improved living conditions and health-care. As the markets for caregivers may not be able to keep up with the unending demand, such development begs the question, what will be the availability of quality care for the elderly in future? The importance of technology in today's world does not escape the impact on the elderly. Falls are a serious health risk among the elderly and often lead to hospitalization [2]. Both extrinsic and intrinsic variables can contribute to falls in older persons [3]. Reduced strength, balance, and flexibility, as well as eyesight impairment, long-term diseases, and adverse drug reactions, are examples of intrinsic variables that are related to the normal aging process. Certain medical conditions, including epilepsy, Parkinson's disease [4], heart attacks, and strokes [5], can also result in falls. Environmental risks include dim illumination; slick surfaces, congested walkways, uneven flooring, and a lack of assistive technology are examples of extrinsic variables. Early diagnosis of falls is crucial, as delaying in medical help can have serious implications, including death. According to a statistics, if a person falls and lies on the floor for a prolonged duration of time, their death risk increases significantly [6, 7]. As a result, timely fall detection, prediction, and alarms are critical. Automatic systems that monitor and report falls can provide useful information for determining causes and establishing prevention strategies. For fall detection, a variety of technologies can be used, such as vision-based systems that use various camera types, such as RGB, infrared, depth, and 3D camera arrays, and wearable sensors like gyroscopes and accelerometers. Figure 1.1 depicts the factors behind elderly fall incidents.

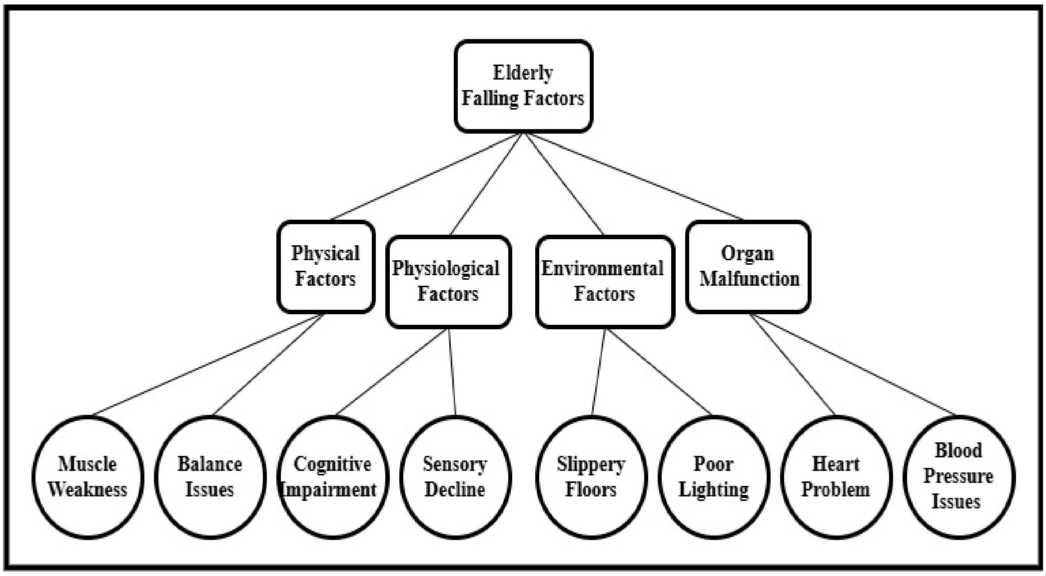

Fig.1.1 Factors behind Elderly Fall Incidents.

Wearable technology has limitations even though it can identify falls by evaluating information such as abrupt changes in angle and speed. It can be difficult for older folks to manage the regular charging requirements of these devices. Wearables are less appropriate for this group since they can also be uncomfortable and have additional negative consequences [8]. Camera-based vision technologies present a viable substitute [9]. They use the capability of contemporary Artificial Intelligence (AI), especially DL [10], which is particularly good at jobs like fall detection, and are non-intrusive and reasonably priced. Vision-based fall detection is a future-proof solution because of the expanding use of Internet of Things (IoT) [11] technologies and the proliferation of cameras in both public and private spaces. Although there are other fall detection techniques, DL is gaining traction because of its better results. DL models is also used for Human Activity Recognition (HAR) like walk, sit, stand, picking-up an item, laying, jump etc. [12]. When it comes to fall detection, DL models can attain great accuracy. Furthermore, methods like few-shot learning and transfer learning make it possible to implement DL models on less potent edge devices, increasing their suitability in a variety of contexts. Figure 1.2 depicts the working principle of vision based fall detection system using DL.

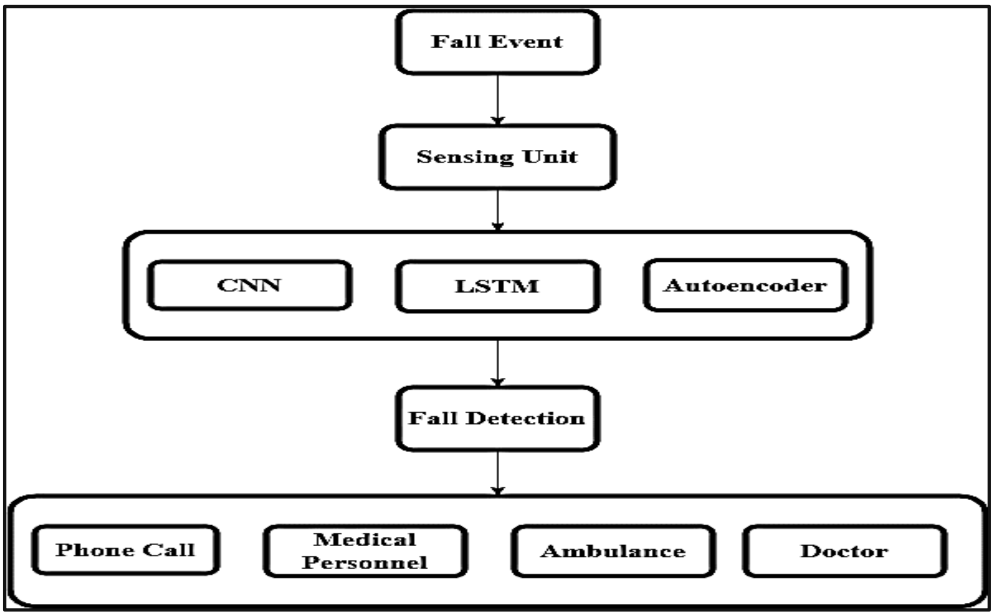

Fig.1.2 Working Principle of Vision based Fall Detection System using DL.

DL is the actually subset of AI and Machine Learning (ML) where Neural Network (NN) layers are utilized for the extraction of feature and the framework talks about the relevance of feature [13]. The epoch of big data has been ushered by advancements in digital technology and its significant usages over internet, and thereby accelerating the progress in DL [14]. Hence, NN can directly learn about the sample images and easily reconstruct the segmented sample images [15].

1.2 Objective

Fall detection in elderly individuals is a critical and widely studied problem due to its direct implications for health and safety. While numerous approaches have been proposed over the years, many rely on RGB frames or wearable sensors, which often face limitations in real-world scenarios such as occlusion, privacy concerns, and environmental noise. Consequently, a universally robust and accurate solution that can be reliably deployed across various environments is still lacking. Therefore, this work aims to develop an efficient and reliable fall detection system by leveraging skeletal pose estimation and DL technique.

The goal of this work is to spot falls in video frames using the skeletal keypoints extracted by Google's MoveNet Thunder model. Unlike analyzing images at pixel level, skeletal-based representation explains human posture and movement by reducing distractions from the background. Yet, several issues continue to exist due to the difficulty of body markers becoming obscured by objects, movements that are similar in different scenarios and complicated motion patterns. All the issues should be solved to ensure that fall detection only happens when someone falls.

These problems are addressed in the work by applying a combined DL approach based on Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks. It extracts spatial information from bone structures and next analyzes the activity movements to separate falls from regular actions. The method proposed here only analyzes skeletal motion over time, not the completely pixel-based scene that makes it more robust and applicable across different environments.

Overall, this work seeks to build a fall detection framework that detects falls accurately and with little human help. The specific aim of the system is to deliver accurate outcomes in places where RGB-based systems are likely to make mistakes. Adopting this approach should greatly increase the dependability and size of fall detection systems. The final part is to analyze the model on widely used datasets such as University of Rzeszow Fall Dataset (URFD) and show a clear explanation of its accuracy, precision, recall, specificity and AUC. The work also studies whether skeletal keypoint-based analysis could be a generalized way to identify human activities in safety critical programs.

1.3 Problem Statement

- **Description:**

Since more individuals are living longer today, supporting their health and safety is a major issue around the world. Falling is a main reason many elderly people get seriously hurt, end up in the hospital, and may face lasting consequences. As more people move to cities and more seniors live by themselves, there is now the need for systems that quickly and correctly recognize falls, so help arrives faster and less harm occurs. Many ways to measure with RGB, video or wearable sensors exist, yet privacy, surrounding settings or the user’s discomfort limit their effectiveness. That is why we require a fall detection system that uses advanced computer vision, is reliable, non-intrusive and operates in real-time. A system for fall detection separates fall events from everyday activities by processing visual or sensor observations. The use of skeletal pose estimation allows the system to retain people’s privacy and deliver quick, easy to read motion information. Monitoring systems that automatically report falls perform much better than human guardians do, especially in hospitals, homes or assisted living centers for the elderly.

- **Challenges:**

1. Traditional video-based or wearable sensor methods are either intrusive, less accurate in complex environments, or raise privacy concerns, especially in indoor settings.
2. Skeletal keypoint estimation, although abstract and privacy preserving, may suffer from occlusion, missing joints, or incorrect poses, which can lead to inaccurate fall classification.
3. Temporal modeling of human motion is complex, simple models often fail to capture subtle differences between falls and fall-like activities (e.g., sitting, bending).
4. The availability of large, diverse, and annotated datasets for fall detection using skeletal keypoints is limited. Current datasets like URFD offer a narrow range of scenarios.
5. Deploying a real-time fall detection system requires optimized models capable of running efficiently on edge devices or within video surveillance systems immediately.

- **Scope:**

Fall detection using skeletal keypoints has immense potential in improving elderly safety and automating critical healthcare alerts. Some important application areas include:

1. Real time monitoring of elderly individuals in homes, hospitals, and care centers.
2. Integration with surveillance systems in public spaces for immediate fall response.
3. Reducing human monitoring efforts while maintaining privacy through pose-based rather than video-based analysis.
4. Providing foundational architecture for broader human activity recognition tasks in safety, fitness, or behavioural analysis domains.

1.4 Motivation

This work provides motivation to improve fall detection systems for elderly individuals in increasingly independent and technology assisted living environments. Today, AI driven healthcare solutions offer a significant advantage in ensuring timely response to critical events like falls, which can lead to severe injuries or fatalities if left undetected. This work integrates MoveNet, which enables accurate real time extraction of human keypoints while preserving privacy, making it a practical alternative to traditional camera based methods. The hybrid use of CNN and BiLSTM models enhances the system's ability to distinguish falls from normal activities, even in complex environments. This approach contributes to the growing shift toward intelligent, non-intrusive elderly care systems, offering reliable monitoring and quicker emergency interventions.

1.5 Thesis Organization

The chapter 1 offers background of fall detection, highlighting its importance in elderly care, supported by relevant statistics. It also defines the key objectives of the research. The chapter 2 gives the details of literature survey of recent practices and prevailing revisions. The chapter 3 gives fundamentals of basics approaches of deep learning, followed by the proposed methodology in Chapter 4. The chapter 5 gives the details of datasets used for fall detection. Chapter 6 discuss about the experiential results and analysis. Finally, the chapter 7 discuss about the conclusion with future scope followed by references.

CHAPTER 2

LITERATURE SURVEY

Fall detection technology has significantly advanced from its initial stages, moving beyond traditional video analysis to more sophisticated and intelligent systems. Vision based fall detection techniques can be categorized into two primary types:

(i) Image processing-based fall detection,

(ii) Human skeleton pose estimation-based fall detection.

When using image processing based methods, professionals analyze raw video or images instead of the skeleton. Steps such as breaking the scene into people, fusing the images and tracking motion allow the software to be used in low light when identifying a fall has proved challenging. Alternatively, technologies that estimate pose using images have become strong contenders for their robustness to distractions like changes in light or noise levels. The methods find important joints on the body and use motion analysis to notice when a fall occurs. Thanks to modern progress, human detection, tracking and keypoint extraction with time analysis now combine leading to light, rapid and effective fall detection.

2.1 Related Work

Ramirez et al. [16] introduced method using human pose recognition on camera images to extract distinctive features in 2021. Such an approach makes it easy to locate the main character in a scene, while ignoring everyone else. The model achieved top results when tested on the UP-Fall dataset using Random Forest (RF), Multilayer Perceptron (MLP), K-Nearest Neighbour (K-NN) and Support Vector Machine (SVM), outshining techniques that use CNNs. More impressively, the RF algorithm performed with 99.51% accuracy. Then, Chhetri et al. [17] put forward a method that depended on dynamic optical flow and rank pooling to record how falls evolve with time. Visual Geometry Group (VGG16) helped the approach attain accuracy gains of approximately 3% and reduced processing time by 40-50 milliseconds compared to existing approaches, on University of Rzeszow fall dataset (URFD), Multicam and Le2i fall detection dataset (FDD). In 2022, Salimi et al. [18] presented a new idea using the Fast

Pose Detection (FPD) method for activity recognition. To assess and train models, the URFD dataset was used and the 1D-CNN and TD-CNN-LSTM achieved fall detection accuracy rates of 98% and 97%. Later that year, Liu et al. [19] introduced a fall detection model for seafarer, leveraging the Blazepose-LSTM architecture. Evaluated on URFD, FDD, and a self-developed seafarer falls dataset, the model achieved 100% accuracy in detecting falls and a 97.95% recognition rate for non-fall activities, operating at a detection frame rate of 29 frames per second. In the same year, Anitha and Priya [20] presented a model called VEFED-DL, which makes the use of DL for a good classification of fall and non-fall events. Its performance is analysed using the Multicam and URFD datasets. The VEFED-DL consistently maintained the highest level of accuracy while considering different parameters. Comparative analysis with some of the latest fall detection methods has shown that VEFED-DL is better than many of those methods in the Multicam dataset, outperforming existing techniques in terms of sensitivity and specificity.

In 2023, Alanazi et al. [21] presented a vision-based system employing a multi-stage approach, including human segmentation, image fusion, and a 3D multi-stream CNN model termed 4S-3DCNN. Evaluated on the Le2i FDD, the system achieved impressive performance metrics, including an accuracy of 99.44%, sensitivity of 99.12%, specificity of 99.12%, and precision of 99.59%. Later in that year, Zi et al. [22] introduced a method integrating a dual illumination estimation technique with the You Only Look Once (YOLOv7) object detection algorithm and DeepSORT tracking to enhance fall detection performance in low-light environments. Evaluated on Le2i FDD and URFD, this method demonstrated improved performance compared to existing state-of-the-art fall detection methods.

Durga Bhavani and Ferni Ukrit [23] created INDCNN-FDC in 2024 to apply an inception-v3 network and efficient pre-processing filters that both improve the quality of images and lessen noise. According to test results on URFD, INDCNN-FDC showed a high performance in fall detection and accuracy of 97.66%, which is better than many existing approaches. Later in the year, Sitpasert et al. [24] designed a method that uses YOLOv5 to detect people, DeepSORT to follow them, Blazepose to find their skeletal features and a CNN-LSTM network to classify falls. The performance of the method was assessed with ImVia Fall, URFD and FallAllD datasets, reaching an average accuracy of 96.66%, sensitivity of 89.95% and specificity of 96.72%. In that same year, Liu and Yow [25] developed a system using ResNet and LSTM. ResNet identifies special features in each frame, whereas LSTM records patterns and links between various frames. Such a combination makes it possible to correctly spot and study how fall events sequence over time. With both URFD and FDD datasets, the approach managed 100% and 97.3% accuracy, exceeding the performance of other latest algorithms. Jain et al. [26] introduced LapseNet the following year, which earned its excellence through high accuracy and broad applicability. All URFD, Multicam, CAUCAFall and UBFC fall experiments showed that LapseNet achieved 100% accuracy, AUC, precision and recall, equally during training and testing.

Cai et al. [27] presented VPE-ViT-FD in 2025, applying a vision transformer and including an IMEM module and an LESA module in the algorithm. Whereas IMEM monitors small motion changes in each frame, LESA looks at local traits appearing in front of the camera. Performance of VPE-ViT-FD improved when evaluated using Le2i and UR datasets. Ablation experiments confirm that both IMEM and LESA help to boost the effectiveness of the system. Figure 2.1 shows the distribution of accuracies of the proposed models in the related work. Table 2.1 shows the taxonomy of all related worked based on vision based fall detection techniques.

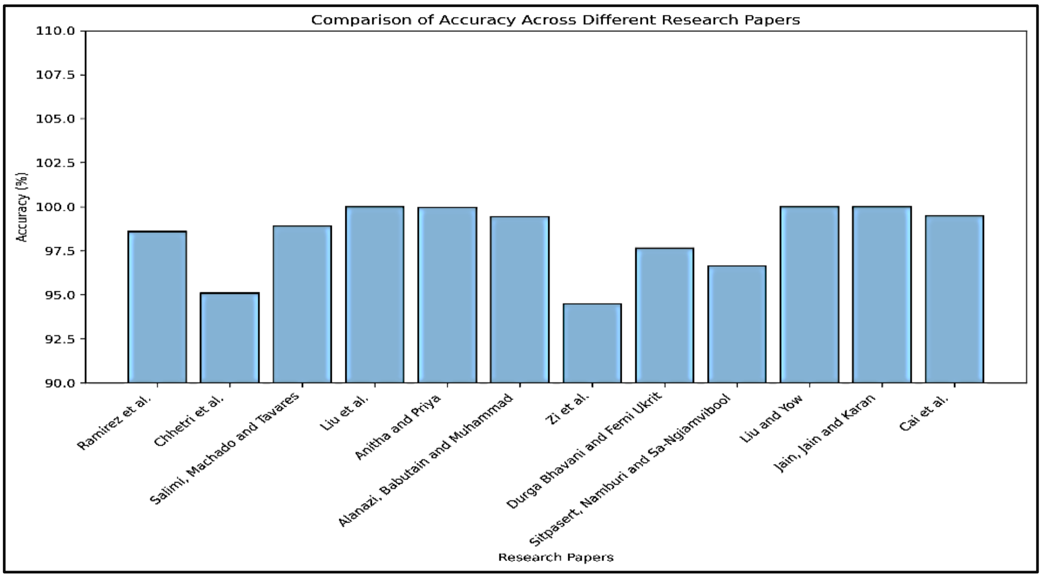

Fig.2.1 Distribution of Accuracies of Proposed Models in Related Work.

Table 2.1 Taxonomy for Vision based Fall Detection Techniques.

Author, Year	**Dataset Used**	**Method**	**Accuracy**	**Pros**	**Cons**
Ramirez *et al.* (2021) [16]	UP Fall	RF, SVM, MLP, KNN, and AlphaPose	98.59%	Accurate, interpretable, multi-person.	No tracking, slower processing.
Chhetri *et al.* (2021) [17]	URFD, Multicam, Le2i FDD	Enhanced optical flow, rank pooling, and VGG16	95.11% (URFD), 91.11% (FDD) and 92.91% (Multi cam)	Efficient, accurate in varying light.	May miss subtle fall cues.
Salimi *et al.* (2022) [18]	URFD	FPD pose estimation with TD-CNN-LSTM and 1D-CNN	98.90%	Accurate, convenient, non intrusive.	Wearables: inconvenient, inaccurate.
Liu *et al.* (2022) [19]	URFD, FDD, self-developed Dataset	Vitruvian-guided BlazePose and LSTM	100%	High accuracy prioritizes sensitivity.	May trigger false alarms.
Anitha and Priya (2022) [20]	Multicam, URFD	VEFED-DL (MobileNet, GRU, and GTOA-SAE)	99.98% (URFD) and 99.92% (Multi cam)	High accuracy, reliable, effective.	Environmental factors may influence performance.

Alanazi *et al.* (2023) [21]	Le2i FDD	4-stream 3D CNN	99.44%	Advanced, accurate Fall detection	Limited scope, single dataset tested.
Zi *et al.* (2023) [22]	Le2i FDD, URFD	YOLOv7, DeepSORT, and Retinex	93.2% (URFD) and 94.5% (Le2i)	Enhanced vision, robust tracking, accurate.	Low-light performance may degrade.
Durga Bhavani and Ferni Ukrit (2024) [23]	URFD	GF, GIF, Inception v3, and Adam-optimized DCNN	97.66%	Accurate, efficient, robust, versatile.	Untested in outdoor environments.
Sitpasert *et al.* (2024) [24]	ImVia Fall, URFD, FallAllD	YOLOv5, DeepSORT, BlazePose, CNN-LSTM	96.66%	User Friendly, low cost, less intrusive.	Environment affects accuracy.
Liu and Yow (2024) [25]	URFD, Le2i FDD	ResNet50 and LSTM	100% (URFD) and 97.3% (FDD)	Non Intrusive fall detection.	Limited real-world data may misclassify actions.
Jain *et al.* (2024) [26]	URFD, Multi cam, CAUCA Fall, UBFC Fall	CNN-LSTM with dropout and Time Distributed layers	100%	Accurate, efficient, and robust.	May overfit, limited environments, needs more data.
Cai *et al.* (2025) [27]	Le2i FDD ,URFD	VPE-ViT-FD: Enhanced Vision Transformer with IMEM and LESA.	99.48% (URFD) and 99.25% (Le2i)	Enhanced under standing, accurate, generalizable.	Original model had limited motion, attention focus.

2.2 Summary

The literature review culminates that although fall detection technologies have significantly progressed, current methods may still face challenges when applied in diverse real world environments. Key areas of concern identified include:

- Many studies achieve high accuracy using specific datasets like URFD, Le2i FDD, or Multicam. However, their generalizability to real world scenarios with varied lighting, backgrounds and multiple people remains limited. There is a need for methods validated across diverse datasets and environments.
- Some approaches lack effective tracking or advanced temporal modeling, potentially leading to misclassification in multi person or continuous action settings. Accurate fall detection requires both spatial and temporal consistency, particularly in crowded or cluttered scenes.
- Most literature emphasizes high accuracy rates, often exceeding 95%. However, real world deployment demands a balanced focus on sensitivity, specificity, false positives, and system latency, which are not always reported or optimized.
- Several methods show performance degradation under low light or outdoor conditions. While some studies attempt to enhance vision (e.g., using Retinex or image fusion), a consistent and reliable performance across lighting conditions remains a concern.
- Despite high accuracy, many proposed methods involve computationally intensive architectures, limiting their deployment on edge devices or in real time systems. Lightweight, efficient, and scalable solutions are still lacking in practice.
- Many techniques are tested only in controlled lab settings. Fall detection systems need rigorous validation in real life, unconstrained environments to ensure their practical applicability, especially for elderly care and healthcare systems.

CHAPTER 3

FUNDAMENTALS OF DEEP LEARNING

A complete approach for human fall detection involving pose estimation and DL is described in this work. To get 17 human pose keypoints from video frames, the approach uses MoveNet Thunder from TensorFlow Hub. The motion dynamics are shown through skeletal annotations and the keypoints are grouped by order, based on when they occur. The proposed method uses both CNN and BiLSTM networks to study both the shape changes and movement patterns in falling humans. In this section, the main technique is reviewed, its principles are stated and you will find an overview of the libraries used in solving the problem, including Keras, TensorFlow and others.

3.1 MoveNet

MoveNet from Google is a modern pose estimation tool that performs fast and accurate for fitness coaching and motion analysis used in health settings. With high efficiency, it finds 17 key body parts and runs smoothly on tiny devices and Internet browsers. The MoveNet model is available in two variants; Lightning is for fast performance at 192×192 input and depth multiplier of 1.0, reaching over 50 FPS, while Thunder chooses greater accuracy with 256×256 input and depth multiplier of 1.75 for speeds of over 30 FPS [28]. Figure 3.1 illustrates the design of the MoveNet model.

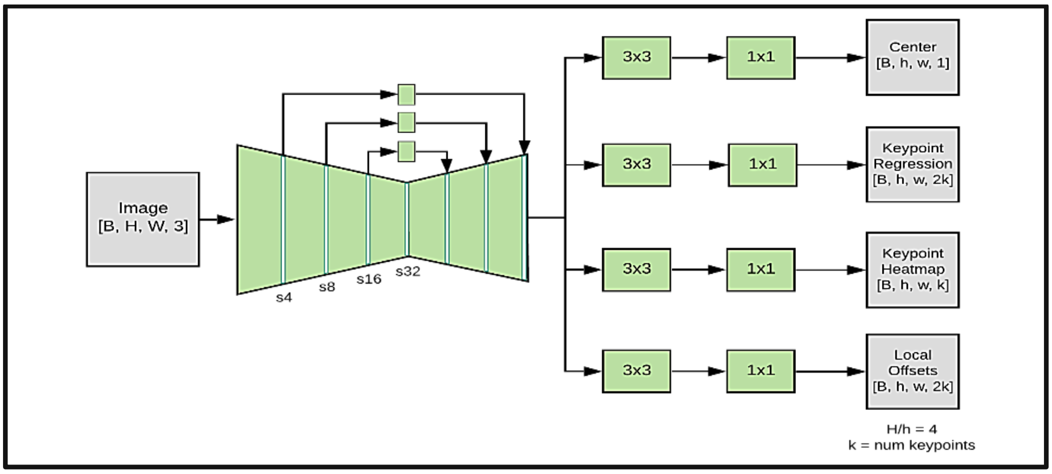

Fig.3.1 Architecture of MoveNet Model [29].

This is achieved through a streamlined architecture that integrates MobileNetV2 as a backbone, enhanced with a Feature Pyramid Network (FPN) and CenterNet prediction heads, ensuring robust detection even under rapid motion in fitness or sports settings [30].

The input to MoveNet is an RGB image tensor, either 192×192×3 for Lightning or 256×256×3 for Thunder, with normalized pixel values. The output structure varies by use case: single-pose estimation returns a (1, 51) tensor containing 17 keypoints as (y, x, confidence) triplets, while multi-pose estimation outputs a (1, 6, 51) tensor to accommodate up to six individuals. Its innovative centre out regression approach begins from detected person centroids using heatmap analysis, followed by keypoint refinement via local offset predictions. This technique eliminates the need for Non-Maximum Suppression (NMS), thereby reducing computational overhead and improving latency, making MoveNet highly suitable for real-time applications that demand both performance and precision [31]. Figure 3.2 depicts the Keypoints generated by MoveNet model.

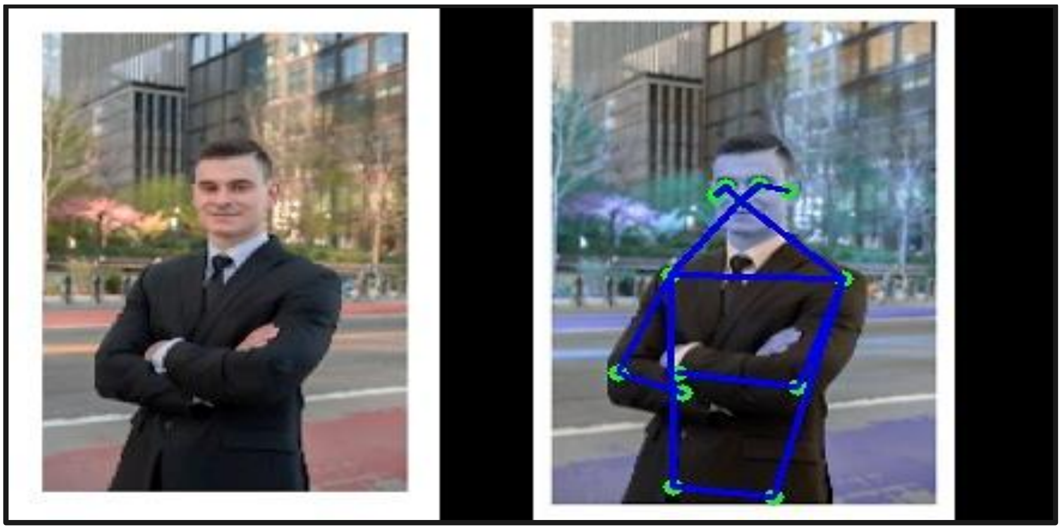

Fig.3.2 Keypoints Generated by MoveNet Model.

3.2 Convolutional Neural Network

CNNs are included in the family of DNNs [32] and are applied in tasks such as validating images and videos in computer vision [33]. This NN architecture is built to work with data arranged in a grid and it does so by running convolutional and pooling operations. A CNN uses convolutional layer, pooling layer and fully connected layer as its basic components. They are in charge of extracting information from data by using a set of filters that glide over the input to produce feature maps. They are used to decrease the size and spacing of the feature maps. To sort the input into categories, well-connected layers make use of the chosen features. To update the filter weights during training, backpropagation computes the changes required in the loss function related to the network's parameters and updates each parameter. Many times through this process, the network learns how to recognize patterns and make accurate predictions. CNNs have excelled on several computer vision tasks such as object detection, image classification and semantic segmentation [33]. They are often useful in fields like NLP [34] and speech recognition [35], where the data is presented as a grid. Alternatively, ML systems take in different instances from a dataset to instruct the chosen machine learning systems. Figure 3.3 illustrates the architecture of CNN model.

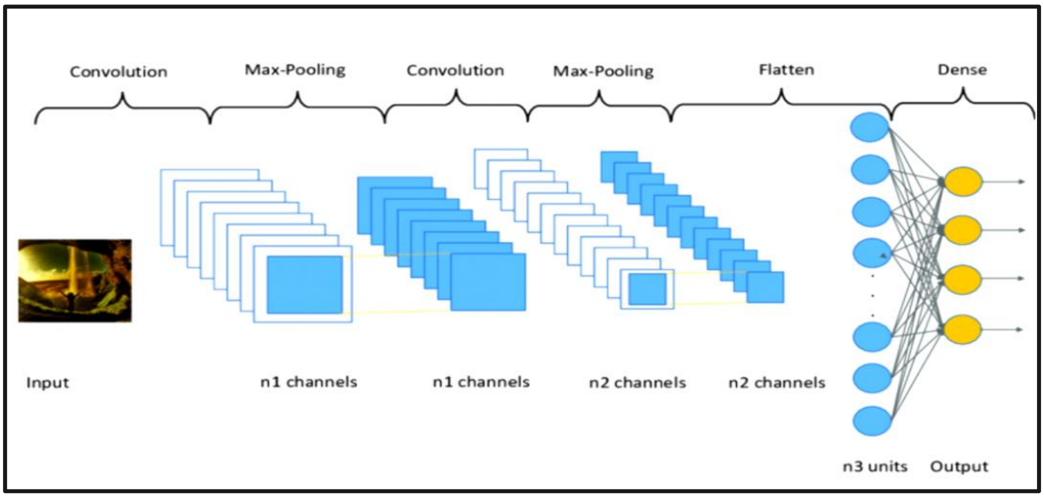

Fig.3.3 Architecture of CNN Model[36].

3.3 Data Acquisition

Data acquisition is the procedure of aggregating data from numerous sources and converting it into a digital format that can be managed via systems or other electronic device. These systems are used in various possible implementations, from scientific research to industrial process control. Its process includes three important key mechanisms:

1. Sensors/devices: These devices calculate physical properties of the environment, like voltage, pressure, and temperature.

2. Data acquisition hardware: This equipment is used to connect the sensors or devices to the system or other electronic machines. Data acquisition hardware involves ADCs that transform the analog signals from the sensors into digital signals and can be managed by a system.
3. Data acquisition software devices: This is the software used to manage the data acquisition hardware and to collect, store, and analyze the data which is being acquired. Data acquisition software can involves drivers for the data acquisition hardware, user interfaces to configuring and manage the hardware, and data processing and analyzing tools.

Data acquisition systems can be fabricated for the widespread series of presentations, from simple data categorization to composite process control and monitoring systems.
Some common uses of data acquisition schemes include:

1. Scientific research: Data acquisition systems are used in scientific research to gather and evaluate data from investigates and clarifications.
2. Industrial process control: Data acquisition systems are used in manufacturing process control to monitor and control developed processes like temperature control in chemical reactions or pressure control in oil drilling.
3. Environmental monitoring: Data acquisition structures are used in environmental monitoring to ration and examine air and water worth, weather conditions, human patterns, and other environmental factors.

In essence, data acquisition helps as a keystone in the gathering and exploration of data from a numerous of sources, allowing enhanced decision-making, process control etc.

3.4 Image Processing

The image processing approaches can apply individually or in combination to attain diverse image processing objectives, ranging from enhancing image quality for visual inspection to extracting quantitative information for further analysis.

There are numerous applications of image processing, like:

1. Medical Images: Image processing is used to analyze medical images, such as X rays, MRI scans, and CT scans, to aid in the diagnosis and treatment of medical conditions.
2. Security-Surveillance: Image processing is used to enhance and analyze

3. surveillance camera images, as well as to identify and track objects or people of interest.
4. Robotics-Automation: Image processing is used in robotics and automation systems to provide vision capabilities, such as object detection and recognition.
5. Entertainment-Education: Image processing is used in the creation of visual effects for movies and video games. Education like results OMR sheets etc.

In recent years, deep learning techniques such as CNNs, RNN have also been used in image processing and pattern recognition and achieving advanced results in various image related tasks, which are, object recognition, segmentation, and image classification.

3.5 Data Augmentation

This is a method of expanding the range of dataset via employing different transformations or modifications to the original datasets. The primary aim behind data augmentation is to enrich the variety of the dataset, eventually enlightening the performance of deep learning models. This improvement is achieved by mitigating overfitting and enhancing the model's ability to generalize to unseen data. The choice of augmentation techniques depends on the specific characteristics of the data and the requirements of the task.

Some common techniques include:
1. Image augmentation: This involves applying transformations such as rotation, flip ping, scaling, cropping, and color adjustments to images.
2. Text augmentation: This involves applying techniques such as synonym replacement, word deletion, and word swapping to text data.
3. Audio augmentation: This involves spread on transformations such as time stretching, pitch shifting, and noise addition to audio data.

Data augmentation can take place manually and it can be analyzed using libraries PyTorch, Keras and TensorFlow from python. In addition, python libraries come with common augmentation functions included.

CHAPTER 4

PROPOSED WORK

This system is intended to accurately study motion patterns by using pose based skeletal keypoints. There are three phases in this model: feature extraction from the skeleton, CNN is used to learn spatial features and BiLSTM networks handle temporal modeling. The final aim is to assign a short sequence of frames to being a fall or non-indicating-of-fall (ADL) event after looking at the extracted posing information. Figure 4.1 depicts the block diagram of the proposed work.

4.1 Human Skeleton Feature Extraction

To ensure robustness and efficiency, we utilize MoveNet Thunder, a lightweight and accurate real-time pose estimation model hosted on TensorFlow Hub. The model is a TFLite version that detects 17 anatomical keypoints for a single person in each frame, including joints such as shoulders, elbows, hips, knees, and ankles.

Each input frame is preprocessed by resizing it to 256×256 pixels and cast to uint8, as required by the model. The model output is a tensor of shape (1, 1, 17, 3) where its normalized coordinates (y, x) represent each keypoint and a confidence score c, given by following equation:

$$\text{Keypoints}_i = (y_i, x_i, c_i), \quad i=1, \ldots, 17 \qquad (1)$$

To form the dataset, consecutive frames are grouped into overlapping temporal windows of 15 frames (timesteps), resulting in 3D tensors of shape (15, 17, 3) encoding the spatial and confidence values of the skeleton over time. Each window is labeled as "fall" or "non-fall" based on the frame wise annotation provided in the URFD dataset CSV files. A utility function maps the ground truth from the annotation file to each frame index in the video.

4.2 CNN for Spatial Feature Learning

Although the pose data is relatively low dimensional compared to raw video frames, spatial patterns between keypoints (e.g., limb orientation, body posture) can be informative for activity classification. To capture these patterns, we apply 1D convolution along the keypoint dimension for each frame in the sequence using TimeDistributed layers. This ensures that the same convolutional operation is applied independently to each frame in the temporal sequence. After convolution and pooling, spatial features from each frame are flattened, resulting in a sequence of feature vectors ready for temporal modeling.

4.3 BiLSTM for Temporal Modeling

Temporal dynamics are essential in distinguishing fall events from other activities. For example, rapid body displacement, collapse, and lack of subsequent movement are strong indicators of a fall. To effectively model these temporal patterns, we employ a Bidirectional LSTM (BiLSTM) layer. The BiLSTM processes the sequence both forward and backward in time, capturing past and future dependencies, and producing a comprehensive feature representation. The final BiLSTM output is passed through a fully connected (dense) layer with 64 ReLU units followed by a dropout layer. The output layer uses a softmax activation to output a two-class probability distribution. Figure 4.2 depicts the structure of the proposed model.

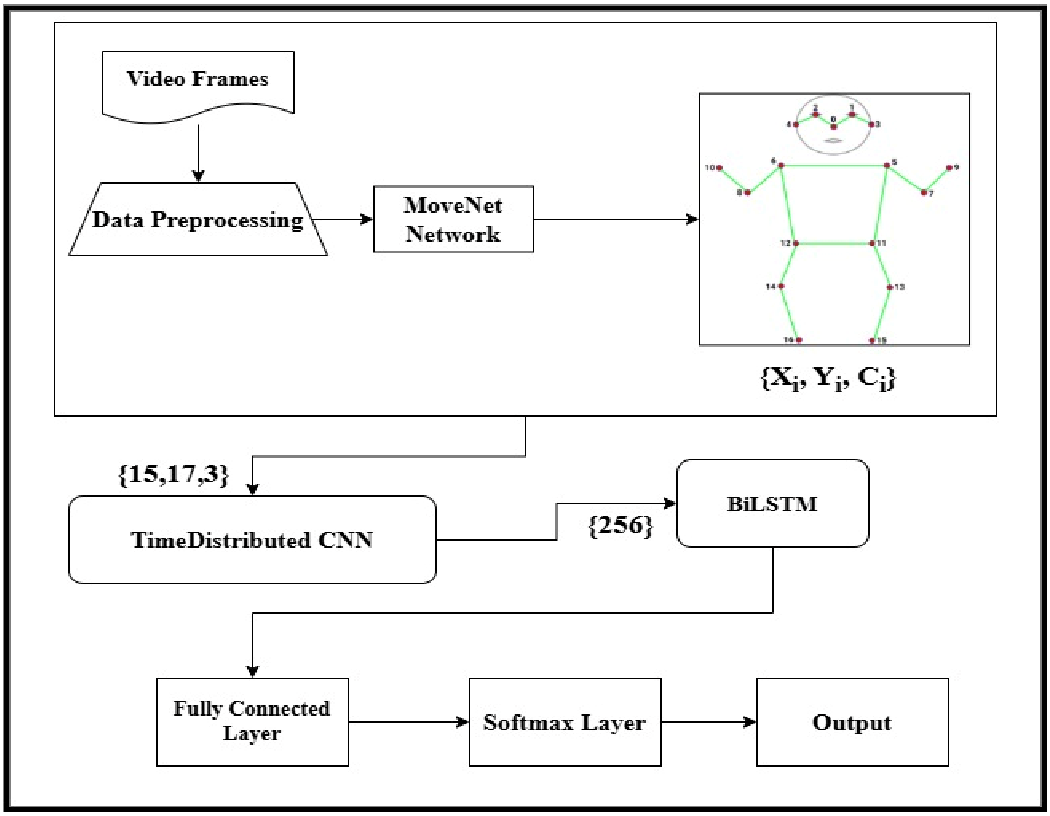

Fig.4.1 Block Diagram of the Proposed Work.

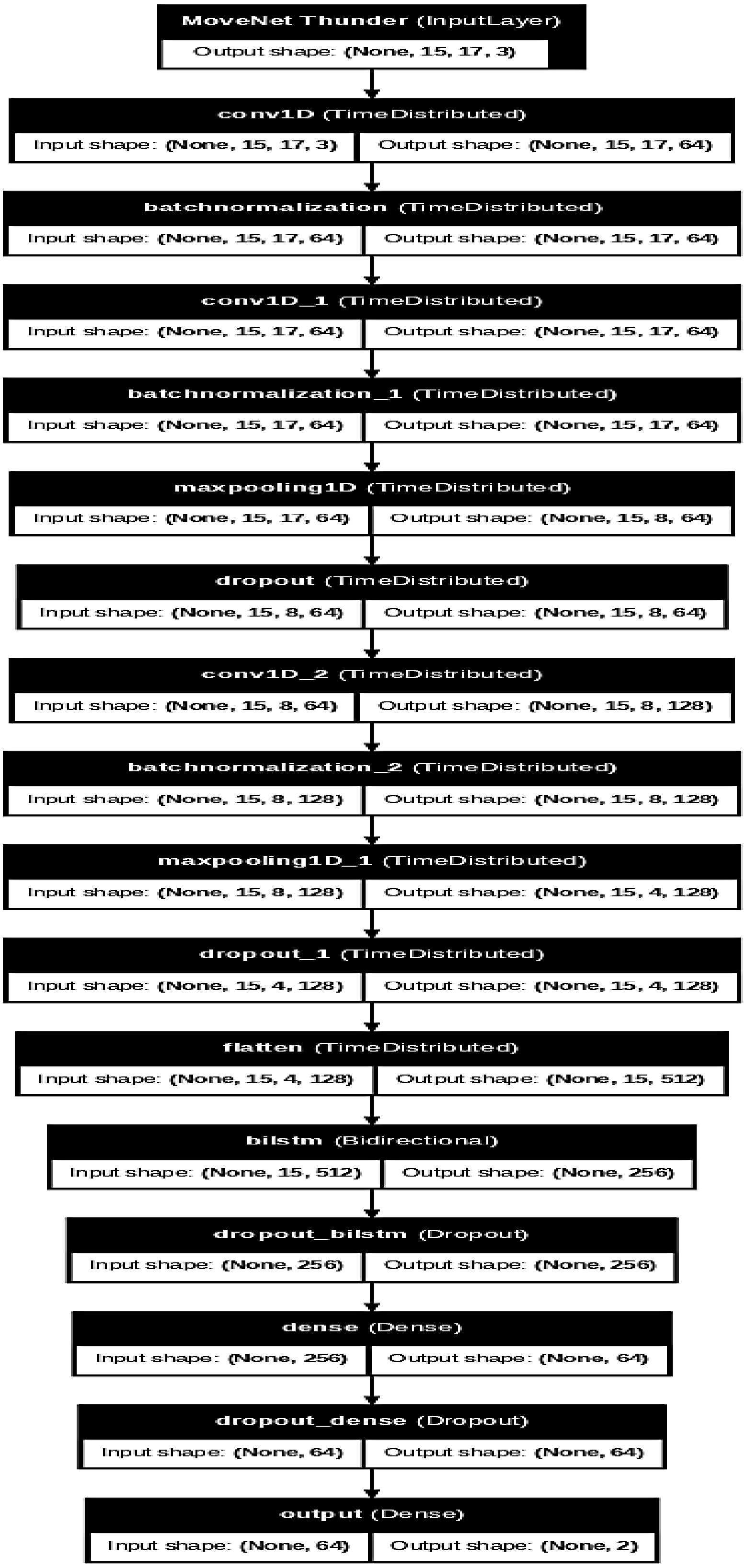

Fig.4.2 Structure of the Proposed Model.

CHAPTER 5

DATASETS

In this research work, URFD dataset [37] is used for the training and testing of the model. We are now going to read the in-depth descriptions of the dataset in the section below.

5.1 URFD Dataset

URFD dataset, created by Kwolek and Kepski in 2014 [37], was recorded using two Microsoft Kinect cameras in indoor environments such as offices and homes, capturing data from five different subjects. The dataset contains 70 RGB-D video sequences, including 30 fall events and 40 non-fall activities (Activities of Daily Living). Each video is annotated with frame-level labels and includes synchronized RGB, depth, skeleton joint data, and IMU signals. In total, the dataset consists of 2,373 fall frames, 7,452 non-fall frames, and 1,719 transition frames. This well labelled and multimodal dataset enables robust training and evaluation of vision-based models. In this work, only the RGB videos and their annotations are used, with pose keypoints extracted for deep learning-based classification. Out of 70 video, we randomly split the dataset into 52 videos for training and 23 videos for validation. Table 5.1 shows the summary of URFD Dataset.

Table 5.1 Summary of URFD Dataset.

Dataset	Camera Type	Fall Type	Data types	# Participants	Environment	# Samples
URFD (2014)	2 Microsoft Kinect	From standing, from sitting on a chair	RGB-D, Skeleton, IMU	5	Office, Home	70

CHAPTER 6

RESULTS AND DISCUSSION

Four landscapes are castoff to establish performance measurements, such as sensitivity, precision, accuracy, and specificity: false positive, true negative, true positive and false negative. This experiment uses performance indicators like sensitivity, specificity, precision, accuracy and area under curve (AUC) to evaluate the classifier's performance. The evaluation measures are defined as:

1. Precision: The percentage of genuine positive predictions to the total of false positives and true positives is measured by the precision performance metric.
2. Recall: Analyzing a model's capacity to prevent false negatives is essential. A high recall score indicates that there is less chance of false negatives because the model is good at identifying a significant percentage of pertinent positive cases.

6.1 Experimental System Setup

This section defines the implementation detail of whole setup. The model has been executed on one fall detection dataset which is publicly available i.e., URFD dataset [37]. The model was executed in a Google Colaboratory environment with 83.5 GB of RAM and an NVIDIA A100 GPU, which significantly accelerated training through high computational throughput. Training was conducted over a maximum of 100 epochs with an initial learning rate of 0.0005 and a batch size of 32. The model utilized the Adam optimizer, along with a learning rate scheduler (ReduceLROnPlateau) and early stopping to prevent overfitting. The best performance was achieved at epoch 21, reaching a training accuracy of 98.85%, a validation accuracy of 96.57%, and a validation loss of 0.0559. The learning rate was progressively reduced during training, and the model weights were restored from the best epoch. The complete training process took approximately 170 seconds, highlighting the model's computational efficiency and fast convergence. Figure 6.1 shows the training and validation results of proposed model on URFD Dataset.

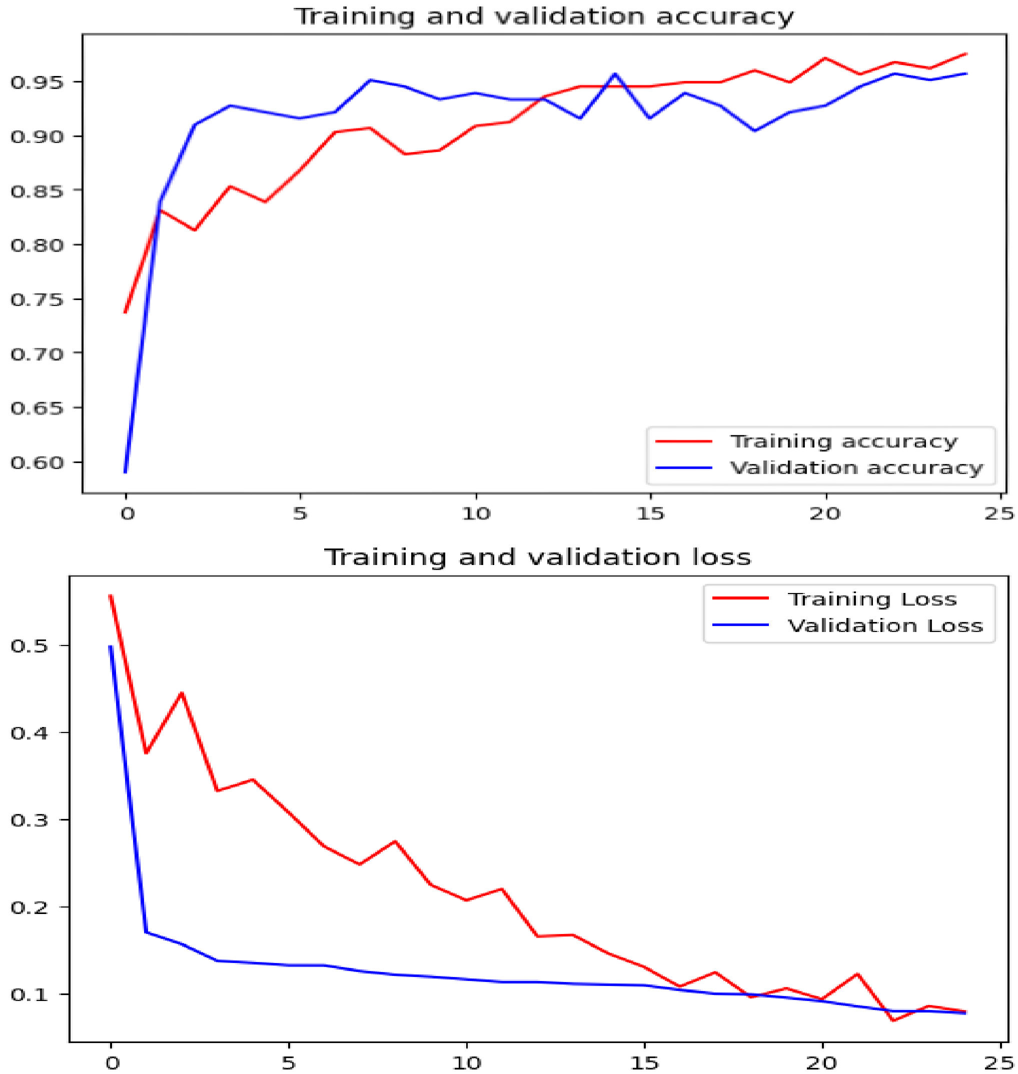

Fig.6.1 Training vs Validation Result on URFD Dataset.

6.2 Methods for Evaluation

Accuracy, Precision, Recall, Specificity, and AUC are the five performance measures utilized in this case to compare the outcomes. These metrics are employed to assess the effectiveness of classification algorithms.

6.2.1 Precision

Precision is the fraction of relevant instances among all retrieved instances. It can be calculated by dividing the number of true positives (TP) by the number of predicted positives (TP + FP). Precision measures how well a model can avoid false positives, or how accurate its positive predictions are. Let us say a doctor wants to determine whether a patient has a certain disease, and uses a diagnostic test to make the determination. The test results can be positive or negative. The doctor performs the test on 100 patients, and the test results show 40 positive results and 60 negative results. The doctor knows from previous experience that the true prevalence of the disease in the patient population is 20%.The doctor is interested in the precision of the test, which refers to the proportion of positive test results that are truly positive. If the doctor examines the 40 patients who tested positive and finds that 30of them truly have the disease, while 10 do not, then the precision of the test is:

Precision = 30 / (30 + 10)

Precision = 0.75 or 75%

This means that the precision of the test is 75%, meaning that 75% of the patients who tested positive actually have the disease, while 25% of the positive results were false positives.

6.2.2 Accuracy

A measurement, computation, or forecast is accurate or precise is referred to as accuracy. The number of accurate forecasts or measurements divided by the overall count of accurate predictions or measurements is often stated as a percentage or ratio. In other words, accuracy assesses a model or systems performance in terms of the accuracy with which it can recognize or categorize data. Accuracy represents the proportion of accurate predictions made by an algorithm among all the predictions it has made. It can be calculated by dividing the number of true positives (TP) and true negatives (TN) by the total number of instances (TP + TN + FP + FN), where FP is false positives and FN is false negatives. Accuracy measures how well a model can classify all instances correctly, regardless of their class. Let us say a company is trying to predict which job candidates will be successful in their role. They use a test to evaluate candidates’ skills, and they use the results of the test to make their hiring decisions. The company hires 100 candidates based on their test results. After six months on the job, the company evaluates how well each employee is performing and categorizes them as either successful or not successful based on predetermined criteria. If the company correctly identified 80 out of the 100 successful candidates using the test, and correctly identified 10 out of the 100 unsuccessful candidates, then the accuracy of their test is:

Accuracy = (No. of correct predictions) / (total no. of predictions)

Accuracy = (80 + 90) / 200

Accuracy = 0.85 or 85%

This means that the test had an accuracy of 85%, meaning that it correctly identified 85% of the candidates who would be successful on the job, and incorrectly identified 15% of the candidates who would not be successful on the job.

6.2.3 Recall

Recall is the fraction of relevant instances that were retrieved. It can be evaluated by dividing the number of true positives (TP) by the number of actual positives (TP + FN)/2. Recall measures how well a model can capture positive cases, or how sensitive it is to positive instances. Let's say a company wants to predict which customers are likely to churn (i.e., stop using their services). They use a machine learning model to make these predictions, which outputs a score for each customer indicating their likelihood of churning. The company has 1,000 customers, of which 200 have already churned. The machine learning model predicts that 300 customers are likely to churn in the future. The company is interested in the recall of the model, which refers to the proportion of actual churners that are correctly identified by the model (i.e., the proportion of true positives among all actual positives).

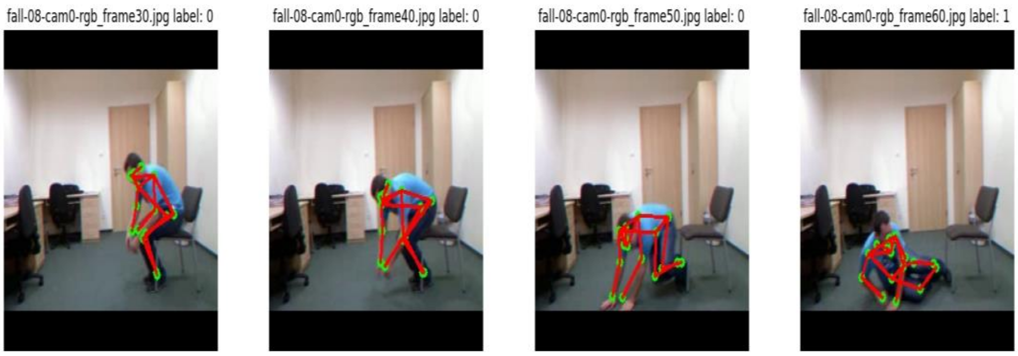

Fig.6.2 Extracted Skeletal Keypoints of frames in FALL sequence along with their Label.

Fig.6.3 Extracted Skeletal Keypoints of frames in ADL sequence along with their Label.

6.3 Results

This subsection shows the comparative results of various methods on the URFD dataset. Table 6.1 listed the result analysis for µAccuracy, µPrecision, and µRecall and µSpecificity of various models, which shows that the proposed model out-perform better results.

Table 6.1. Comparative Results of different methods for URFD Dataset.

Model	**µAccuracy (%)**	**µPrecision (%)**	**µRecall (%)**	**µSpecificity (%)**
CNN[38]	95.1	71.8	71.3	99.5
LSTM[39]	92.4	94	83.9	-
GRU[39]	85.2	93.7	76.7	-
CNN+LSTM[40]	98.59	91.08	94.37	98.96
Proposed(Ours)	95.80	93.90	93.75	94.78

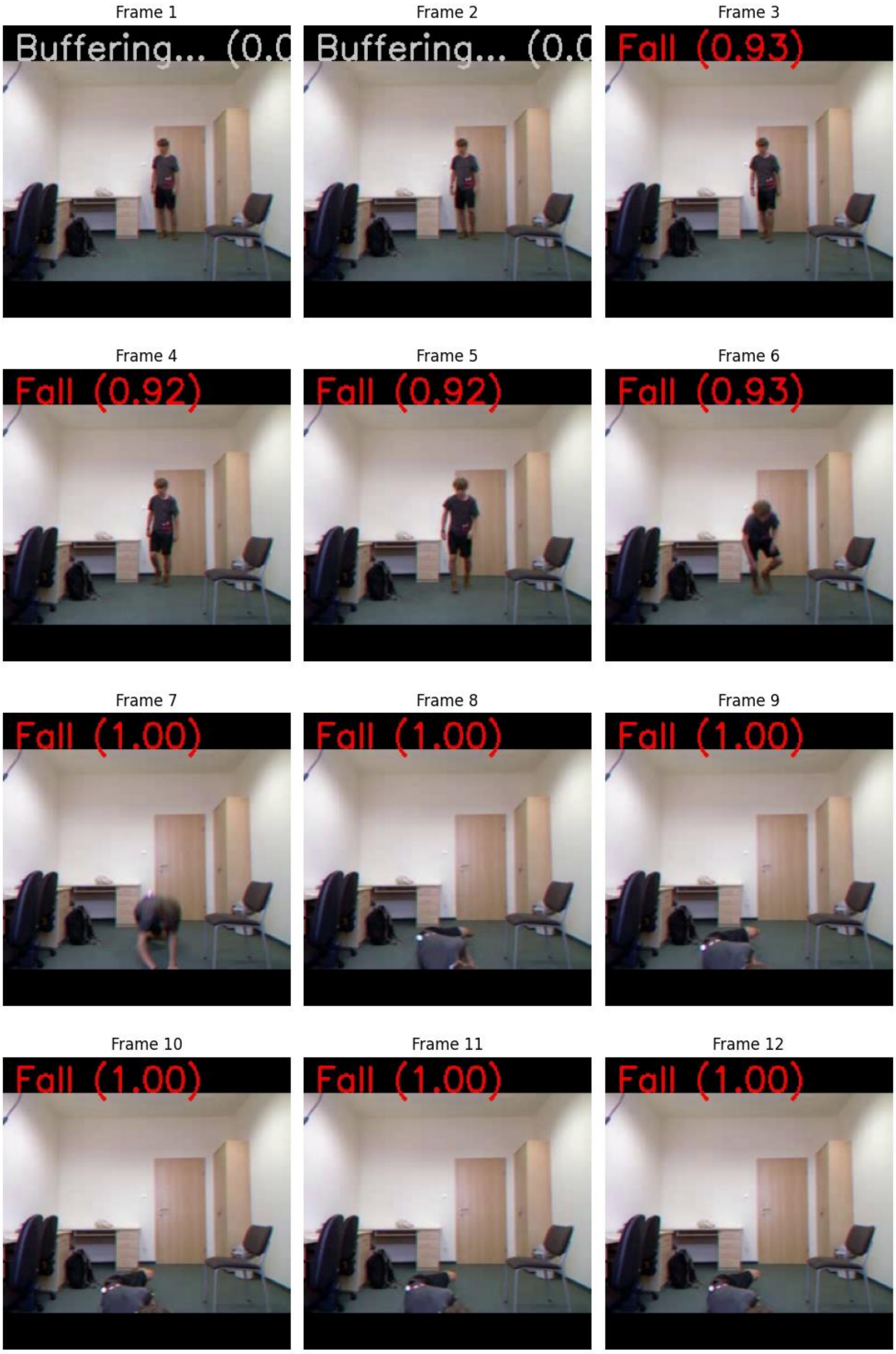

Fig.6.4 Result obtained for the fall sequence using proposed model.

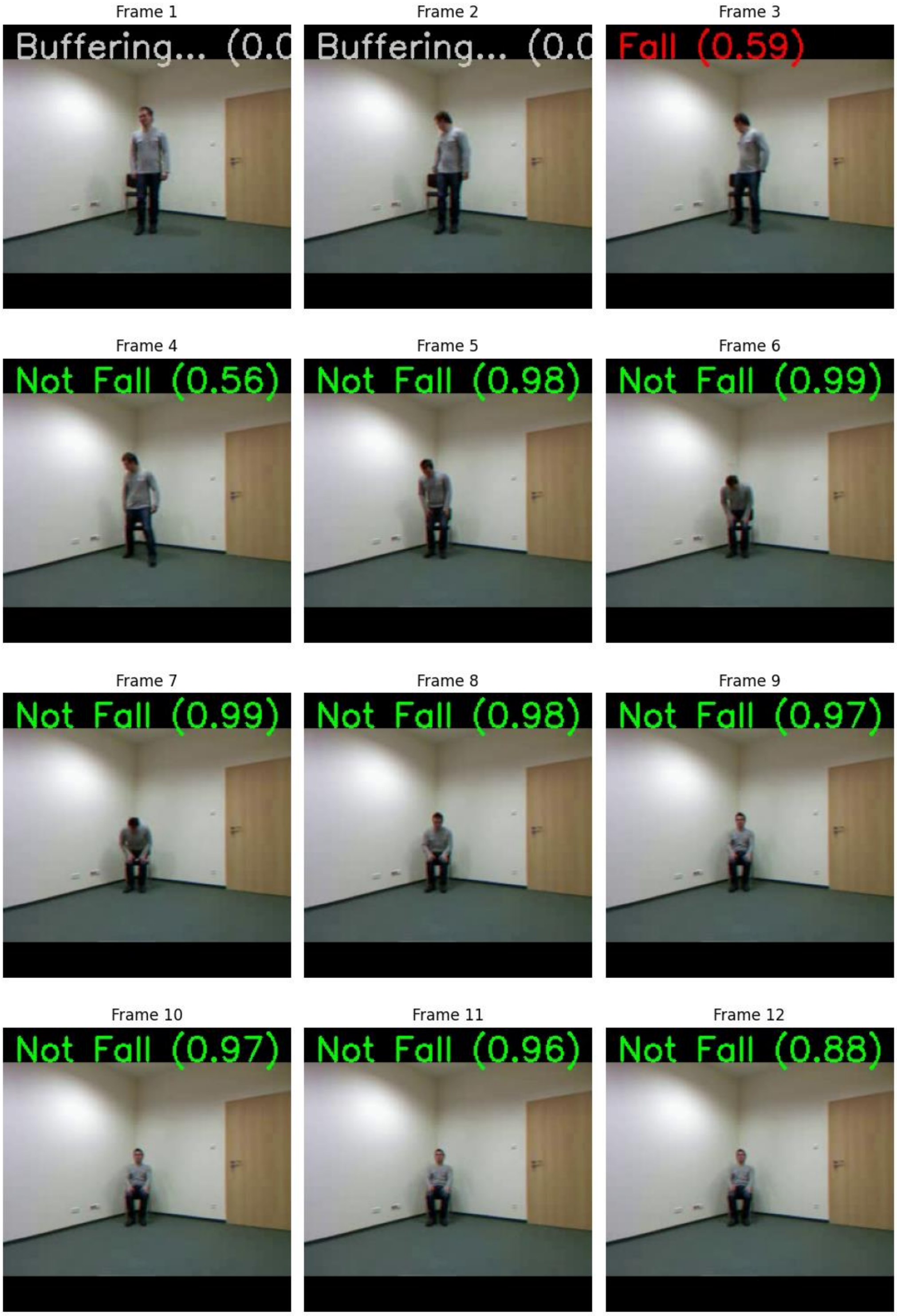

Fig.6.5 Result obtained for the adl sequence using proposed model.

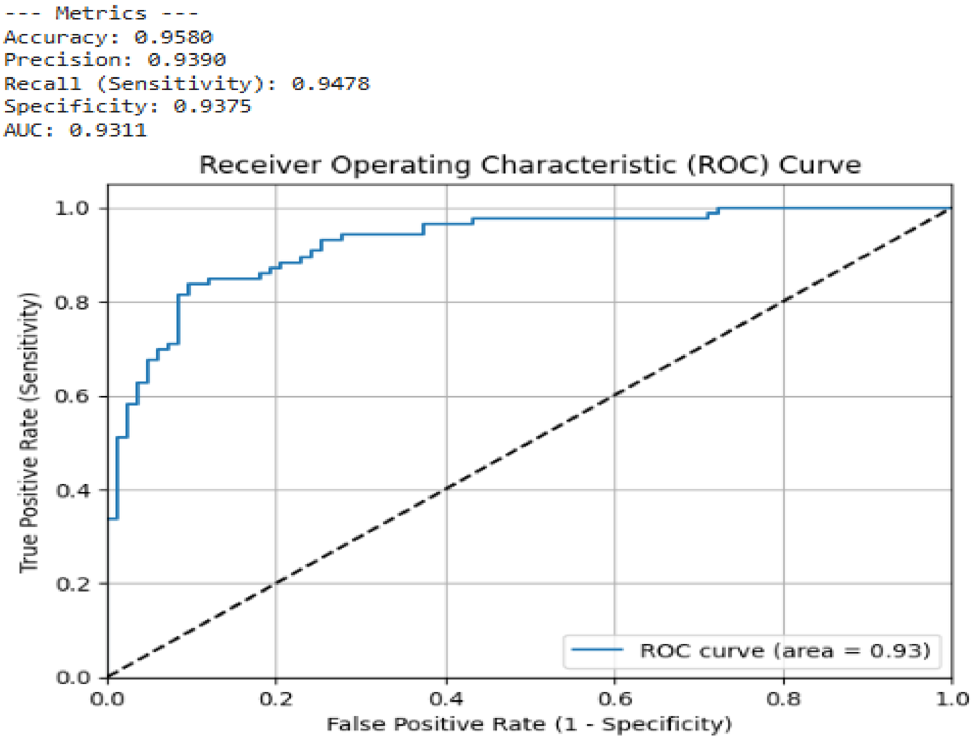

Fig.6.6 Results of Experimental Evaluation.

CHAPTER 7

CONCLUSION AND FUTURE WORK

We designed a system in this study that detects falls by monitoring movements with MoveNet Thunder, then using a combination of CNN and BiLSTM to tell falls apart from ADLs. We process RGB videos, detect, and display 17 important human body joints, forming sequences to study movement over time. Testing and analysis were performed on the URFD dataset and the model showed good performance, reaching 95.80% accuracy, 93.90% precision, 94.78% recall, 93.75% specificity and a 0.9311 AUC. The system reliably detects falls while causing very few false alarms, so it can safely be deployed in places dealing with the elderly's safety.

However, this work also has some limitations despite its effectiveness. Because it was annotated properly despite having only a limited number of people and indoor falls, the URFD dataset was the one we used to develop and check the model. The result is that the model does not do as well dealing with different lighting and angles. Only information from RGB images was used because depth and IMU data were not part of the processing. Still, using pose information speeds up the process and also improves security, but small issues with body posture can cause accuracy to drop in poorly lit or concealed areas.

Making the model work better means uniting pose information with audio and wearable data through a single multimodal setup. If data from different settings, different subjects and actual falls were used, the robot would work better in a variety of situations. To make the model even better, we can use it in web applications that control many cameras, suitable for using in homes, hospitals and elderly care centers. In addition, adding localization features to video can help pinpoint the moment of the tumble. The aim of the improvements is to make the system a solid and flexible tool for serious fall prevention and care.

REFERENCES

[1] United Nations, Department of Economic and Social Affairs, "World population ageing: Ten key messages," Policy Brief No. 10, 2020. [Online]. Available: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Sep/un_pop_2020_pf_ageing_10_key_messages.pdf.

[2] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Monocular 3D head tracking to detect falls of elderly people," in *Proc. 2006 Int. Conf. IEEE Eng. Med. Biol. Soc.*, New York, NY, USA, 2006, pp. 6384-6387.

[3] P. Iamtrakul, S. Chayphong, S. Jomnonkwao, and V. Ratanavaraha, "The association of falls risk in older adults and their living environment: A case study of rural area, Thailand," *Sustainability*, vol. 13, no. 24, p. 13756, 2021.

[4] N. Mühlenfeld *et al.*, "Seizure related injuries - Frequent injury patterns, hospitalization and therapeutic aspects," *Chin. J. Traumatol.*, vol. 25, no. 5, pp. 272–276, 2022.

[5] S. Salavatidezfouli, A. Alizadeh, M. Barzegar Gerdroodbary, *et al.*, "Investigation of the stent induced deformation on hemodynamic of internal carotid aneurysms by computational fluid dynamics," *Sci. Rep.*, vol. 13, p. 7155, 2023.

[6] D. Wild, U. S. Nayak, and B. Isaacs, "How dangerous are falls in old people at home?," *Br. Med. J. (Clin. Res. Ed.)*, vol. 282, no. 6260, pp. 266–268, 1981.

[7] F. Marquis-Faulkes, S. Mckenna, P. Gregor, and A. Newell, "Gathering the requirements for a fall monitor using drama and video with older people," *Technol. Disabil.*, vol. 17, 2005.

[8] A. Sufian, C. You, and M. Dong, "A deep transfer learning-based edge computing method for home health monitoring," *arXiv preprint*, 2021.

[9] R. Espinosa *et al.*, "A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset," *Comput. Biol. Med.*, vol. 115, p. 103520, 2019.

[10] E. Alam, A. Sufian, A. Das, A. Bhattacharya, M. Ali, and M. M. Rahman, "Leveraging deep learning for computer vision: A review," 2021.

[11] A. Sufian *et al.*, "Deep learning in computer vision through mobile edge computing for IoT," in *Proc. Int. Conf. Commun. Signal Process.*, 2021.

[12] A. Sharma and R. Singh, "ConvST-LSTM-Net: Convolutional Spatiotemporal LSTM Networks for Skeleton-Based Human Action Recognition," *Int. J. Multimed. Inf. Retr.*, vol. 12, no. 2, p. 34, 2023.

[13] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

[14] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[15] Hadji, I., & Wildes, R. P. (2018). What do we understand about convolutional networks?. *arXiv preprint arXiv:1803.08834*.

[16] H. Ramirez *et al.*, "Fall detection and activity recognition using human skeleton features," *IEEE Access*, vol. 9, pp. 33532–33542, 2021.

[17] S. Chhetri *et al.*, "Deep learning for vision-based fall detection system: Enhanced optical dynamic flow," *Comput. Intell.*, vol. 37, no. 1, pp. 578–595, 2021.

[18] M. Salimi, J. J. M. Machado, and J. M. R. S. Tavares, "Using deep neural networks for human fall detection based on pose estimation," *Sensors*, vol. 22, no. 12, 2022.

[19] W. Liu *et al.*, "Fall detection for shipboard seafarers based on optimized BlazePose and LSTM," *Sensors*, vol. 22, no. 14, 2022.

[20] G. Anitha and S. B. Priya, "Vision-based real-time monitoring system for elderly fall event detection using deep learning," *Comput. Syst. Sci. Eng.*, vol. 42, no. 1, pp. 87–103, 2022.

[21] T. Alanazi, K. Babutain, and G. Muhammad, "A robust and automated vision-based human fall detection system using 3D multi-stream CNNs with an image fusion technique," *Appl. Sci.*, vol. 13, no. 12, 2023.

[22] X. Zi *et al.*, "Detecting human falls in poor lighting: Object detection and tracking approach for indoor safety," *Electronics*, vol. 12, no. 5, 2023.

[23] K. Durga Bhavani and M. Ferni Ukrit, "Design of inception with deep convolutional neural network-based fall detection and classification model," *Multimed. Tools Appl.*, vol. 83, no. 8, pp. 23799–23817, 2024.

[24] P. Sitpasert, A. Namburi, and W. Sa-Ngiamvibool, "Vision-based fall detection using convolution neural network and long short-term memory," *ICIC Express Lett.*, vol. 18, no. 3, pp. 209–218, 2024.

[25] X. Liu and K. C. Yow, "Human fall detection based on ResNet and LSTM network in surveillance cameras," in *Proc. IEEE Pac. Rim Conf. Commun. Comput. Signal Process. (PACRIM)*, 2024.

[26] S. Jain, S. Jain, and A. Karan, "LapseNet: A hybrid CNN-LSTM approach for accurate and efficient vision-based fall detection," in *Proc. 6th Int. Conf. Robot. Comput. Vis. (ICRCV)*, 2024, pp. 12–17.

[27] X. Cai *et al.*, "Visual perception enhancement fall detection algorithm based on vision transformer," *Signal Image Video Process.*, vol. 19, no. 1, 2025.

[28] MoveNet.SinglePose Model Card, “MoveNet.SinglePose Model Card” [Online]. Available:https://storage.googleapis.com/movenet/MoveNet.SinglePose%20Model%20Card.pdf.

[29] TensorFlow Blog, “Next-Generation Pose Detection with MoveNet and TensorFlow.js” [Online]. Available: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html.

[30] Jo, B., Kim, S. (2022). Comparative analysis of OpenPose, PoseNet, and MoveNet models for pose estimation in mobile devices. Traitement du Signal, Vol. 39, No. 1, pp. 119-124.

[31] TensorFlow Blog, “MoveNet: Ultra fast and accurate pose detection model.” [Online]. Available: https://www.tensorflow.org/hub/tutorials/movenet.

[32] C. M. Bishop, “Neural networks and their applications,” Review of Scientific Instru- ments, vol. 65, no. 6, p. 1803, Jun. 1998.

[33] R. A. Jarvis, “A Perspective on Range Finding Techniques for Computer Vision,” IEEE Trans Pattern Anal Mach Intell, vol. PAMI-5, no. 2, pp. 122–139, 1983.

[34] K. R. Chowdhary, “Natural Language Processing,” Fundamentals of Artificial Intel- ligence, pp. 603–649, 2020.

[35] M. Halle and K. Stevens, “Speech Recognition: A Model and a Program for Research,” IRE Transactions on Information Theory, vol. 8, no. 2, pp. 155–159, 1962.

[36] “Python TensorFlow- tf.keras.layers.Conv2D () Function: GeeksforGeeks.” https://www.geeksforgeeks.org/python-tensorflow-tf-keras-layers-conv2d-function/.

[37] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 489–501, Dec. 2014.

[38] Martínez-Villaseñor, L.; Ponce, H.; Brieva, J.; Moya-Albor, E.;Núñez-Martínez, J.; Peñafort-Asturiano, C. “Up-fall detection dataset: a multimodal approach.” Sensors 19, 1988 (2019)

[39] Lin, C.-B.; Dong, Z.; Kuan, W.-K.; Huang, Y.-F. “A frameworkfor fall detection based on openpose skeleton and lstm/gru models.” Appl. Sci. 11, 329 (2021)

[40] Inturi, Anitha & v. M., Manikandan & Garrapally, Vignesh. (2022). “A Novel Vision-Based Fall Detection Scheme Using Keypoints of Human Skeleton with Long Short-Term Memory Network.” Arabian Journal for Science and Engineering.

LIST OF PUBLICATION(S)

[1] Piyush Agarwal, Abhilasha Sharma, “A Comprehensive Review of Vision-Based Deep Learning Techniques for Fall Detection in Elderly Individuals”. The Paper has been accepted at 2nd International Conference on Engineering, Management, and Social Sciences (ICEMSS-25), 19-20 February 2025. Paper Id: ICEMSS_56.

[2] Piyush Agarwal, Abhilasha Sharma, “Vision Based Elderly Human Fall Detection System using MoveNet Pose Estimation and Hybrid CNN BiLSTM Architecture”. The Paper has been accepted at National Conference on Advanced Computer Science and Information Technology (NCACSI - 25), 1 June 2025. Paper Id: National Conference_9467351.

PROOF OF PUBLICATION

Paper-1

---------- Forwarded message ---------
From: **ICEMSS India** <info@icemss.in>
Date: Fri, Jan 31, 2025 at 2:25 PM
Subject: Acceptance Letter for ICEMSS-25
To: <abhilasha_sharma@dce.ac.in>

Dear Abhilasha Sharma,

Greetings of the day!!!

We are glad to inform you that the article titled **"A Comprehensive Review of Vision-Based Deep Learning Techniques for Fall Detection in Elderly Individuals"** has been accepted for **Virtual Presentation** during the conference **ICEMSS-25, Delhi, India** from **19th-20th February, 2025.**

Here I have attached the acceptance letter for your reference.

Thank you for your interest in ICEMSS-25. To complete your payment registration, please follow the link below: https://www.icemss.in/registration

For more information, Please feel free to contact us.

Regards,
Priyanka Sahu
Program coordinator
ICEMSS - 25
WhatsApp-(+91-8260080050)

ZEP RESEARCH

Zep Research
DCB-330, 3rd Floor, DLF Cyber City, Patia, Bhubaneswar, 751024, IN
+91 8260080050 | www. zepresearch.com |info@zepresearch.com

Online Payment Receipt

BILL TO

Name:	Piyush Agarwal	**Date:**	5/02/2025
Address:	India	**Ref ID:**	ICEMSS-25_56
University:	Delhi Technological University		

Sl No.	Description	Qty	Unit Price	Amount
1	Conference Fee	1	INR 1500	INR 1500
			Total Amount	INR 1500
			Service Tax	
			S & H	
			Discount	
			Grand Total	INR 1500

TAN : BBNC02814A

Amount in Words: ₹ One Thousand Five Hundred Only

ZEP RESEARCH "PIONEERING INNOVATION, SHAPING THE FUTURE"

CERTIFICATE

- OF PRESENTATION -

THIS CERTIFICATE IS PRESENTED TO :

Piyush Agarwal

of *"Delhi Technological University, India"* for presenting his/her research article titled *"A Comprehensive Review of Vision-Based Deep Learning Techniques for Fall Detection in Elderly Individuals"* at the *"2nd International Conference on Engineering, Management and Social Sciences"* held on 19th - 20th Feb 2025 at Indraprastha College for Women-University of Delhi, India, organized by ZEP Research.

Mrs. Priyanka Sahu	Dr. J P Singh	Dr. Alfe M. Solina
Director	Organizing Secretary	Keynote Speaker
Zep Research	ICEMSS-25	ICEMSS-25

Submission Guidelines

All submissions will undergo a double-blind peer review process. Accepted papers will be published in the conference proceedings, with selected papers considered for publication in partner journals. Submissions should highlight the interdisciplinary nature of the work and its potential impact on addressing real-world challenges.

- **Abstract length: 300-500 words**
- **Full paper length: 6-8 pages (including references)**
- **Format: IEEE conference template**

Publication Support in Journals

Paper-2

NATIONAL CONFERENCE

NCACSI - 25

National Conference on Advanced Computer Science and Information Technology (NCACSI - 25)

1st June 2025 Bangalore, India

Acceptance Letter

Authors Name: Mr. Piyush Agarwal, Dr. Abhilasha Sharma

Dear Authors,

We are pleased to inform you that your paper has been accepted by the review committee for Oral / Poster Presentation at the **National Conference on Advanced Computer Science and Information Technology (NCACSI - 25)**

Article Title: **Vision Based Elderly Human Fall Detection System using MoveNet Pose Estimation and Hybrid CNN BiLSTM Architecture**

Paper ID: **National Conference_9467351**

This conference will be held on **1st June 2025 in Bangalore, India**

Your paper will be published in the conference proceeding and Well reputed journal after registration.

Please register as soon as possible in order to secure your participation:

https://www.nationalconference.in/event/registration.php?id=3333157

You are requested to release the payment and mail us the screen of successful payment release with your name and title of paper to confirm your registration.

Sincerely,

NATIONAL CONFERENCE NC

Dr. Tara Srivastava

National Conference

Success

Your payment has been made

Dear Piyush Agarwal,

We are pleased to inform you that your online payment transaction has been successfully processed. The details of your payment are as follows:

Order ID: 682da4b5516f1

Reference ID: 113773896641

Amount: 3600.00 INR

Name: Piyush Agarwal

Email ID: piyushagarwal_23swe10@dtu.ac.in

Phone Number: 91 6387573071

Thank you for choosing **National Conference**
. We appreciate your prompt payment.

For assistance or inquiries, feel free to reach out to us at **info@nationalconference.in** or **+91 9677007228**.

Best Regards,
National Conference
info@nationalconference.in
+91 9677007228

Indexing & Supported By

Google Scholar

CERTIFICATE

OF PRESENTATION

NATIONAL CONFERENCE

National Conference on Advanced Computer Science and Information Technology (NCACSI-25)

01st June 2025 | Bangalore - India

This is to certify that................................**Piyush Agarwal**................................affiliated with

................................Delhi Technological University, India................................has presented a paper titled

Vision Based Elderly Human Fall Detection System using MoveNet Pose Estimation and Hybrid CNN BiLSTM Architecture

at the esteemed Conference organized by the National Conference (NC) held on 01st June 2025at Bangalore - India

CiteFactor
Academic Scientific Journals

Dr. Tara Srivastava
Director

NATIONAL CONFERENCE NC

Raj Kumar
Co-ordinator

Crossref

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis Elderly Individual Fall Detection System Using Deep Learning

Total Pages 30 Name of the Scholar PIYUSH AGARWAL

Supervisor (s)

(1) Dr. Abhilasha Sharma

(2)

(3)

Department Software Engineering

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 7%, Total Word Count: 7024

Date: 22-May-2025

Piyush
Candidate's Signature

Abhilasha Sharma
Signature of Supervisor(s)

Piyush_Agarwal.pdf

Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:97165635

Submission Date

May 22, 2025, 1:18 PM GMT+5:30

Download Date

May 22, 2025, 1:23 PM GMT+5:30

File Name

Piyush_Agarwal.pdf

File Size

1.1 MB

30 Pages

7,024 Words

38,503 Characters

AS

7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 8 words)

Match Groups

44 Not Cited or Quoted 7%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

4%	Internet sources
4%	Publications
4%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **44** Not Cited or Quoted 7%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

4%	Internet sources
4%	Publications
4%	Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	www.mdpi.com	1%
2	Publication	Shrawan Kumar Trivedi, Abhijit Deb Roy, Praveen Kumar, Debashish Jena, Avik Si...	<1%
3	Publication	Mehdi Ghayoumi. "Generative Adversarial Networks in Practice", CRC Press, 2023	<1%
4	Submitted works	Liverpool John Moores University on 2023-03-05	<1%
5	Internet	www.verifiedmarketresearch.com	<1%
6	Submitted works	Glasgow Caledonian University on 2025-05-15	<1%
7	Internet	elifesciences.org	<1%
8	Internet	scholar.archive.org	<1%
9	Internet	www.valuemarketresearch.com	<1%
10	Internet	assets-eu.researchsquare.com	<1%

11 Internet

datarade.ai <1%

12 Internet

www.semanticscholar.org <1%

13 Submitted works

University of Newcastle on 2011-05-16 <1%

14 Internet

www.enggjournals.com <1%

15 Internet

www.politesi.polimi.it <1%

16 Internet

www.hindawi.com <1%

17 Submitted works

Academic Library Consortium on 2024-06-19 <1%

18 Submitted works

Universiti Teknikal Malaysia Melaka on 2016-05-20 <1%

19 Submitted works

University of Birmingham on 2024-11-28 <1%

20 Publication

Khaled M. Abd-Elwahab, Mohamed Rehan, Mohammed A.-M. Salem, Hisham Oth... <1%

21 Internet

fastercapital.com <1%

22 Internet

www.researchgate.net <1%

23 Publication

"Neural Information Processing", Springer Science and Business Media LLC, 2017 <1%

24 Submitted works

Cranfield University on 2012-06-06 <1%

25 Submitted works

De Montfort University on 2013-12-13 <1%

26 Submitted works

Institute of Research & Postgraduate Studies, Universiti Kuala Lumpur on 2024-0... <1%

27 Publication

Mendes, Fábio Barbosa. "Time-Series Analysis Based on Machine Learning for Occ... <1%

28 Publication

Prabhu Rajendran. "Falls, Fall Prevention, and Fall Detection Technologies", Aging... <1%

29 Publication

Rainer Planinc. "Emergency System for Elderly – A Computer Vision Based Approa... <1%

30 Submitted works

Technological Institute of the Philippines on 2025-05-16 <1%

31 Internet

isip.piconepress.com <1%

32 Internet

www.coursehero.com <1%

33 Internet

www.nature.com <1%

Piyush_Agarwal.pdf

Delhi Technological University

Document Details

Submission ID

trn:oid:::27535:97165635

Submission Date

May 22, 2025, 1:18 PM GMT+5:30

Download Date

May 22, 2025, 1:23 PM GMT+5:30

File Name

Piyush_Agarwal.pdf

File Size

1.1 MB

30 Pages

7,024 Words

38,503 Characters

0% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

0 AI-generated only 0%
Likely AI-generated text from a large-language model.

0 AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

DECLARATION

We/I hereby certify that the work which is presented in the Major Project-II/Research Work entitled Elderly Individual Fall Detection Using Deep learning in fulfilment of the requirement for the award of the Degree of Bachelor/Master of Technology in Software Engineering and submitted to the Department of Software Engineering, Delhi Technological University, Delhi is an authentic record of my/our own, carried out during a period from July to Dec 24 under the supervision of Dr. Abhilasha Sharma

The matter presented in this report/thesis has not been submitted by us/me for the award of any other degree of this or any other Institute/University. The work has been published/accepted/communicated in SCI/ SCI expanded/SSCI/Scopus indexed journal OR peer reviewed Scopus indexed conference with the following details:

Title of the Paper: A Comprehensive Review of Vision- Based Deep learning Technique for Fall Detection in Elderly Individual
Author names (in sequence as per research paper): PIYUSH AGARWAL, Dr. Abhilasha Sharma
Name of Conference/Journal: ICEMSS-2025
Conference Dates with venue (if applicable): 19- 20th Feb
Have you registered for the conference (Yes/No)?: yes
Status of paper (Accepted/Published/Communicated): Accepted
Date of paper communication: 30th Jan 2025
Date of paper acceptance: 31st Jan 2025
Date of paper publication: Yet to be done

23/SWE/10
PIYUSH AGARWAL

Student(s) Roll No., Name and Signature

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I, further certify that the publication and indexing information given by the students is correct.

Supervisor Name and Signature

Place: Delhi

Date: 26/05/25

NOTE: PLEASE ENCLOSE RESEARCH PAPER ACCEPTANCE/ PUBLICATION/COMMUNICATION PROOF ALONG WITH SCOPUS INDEXING PROOF (Conference Website OR Science Direct in case of Journal Publication).

DECLARATION

We/I hereby certify that the work which is presented in the Major Project-II/Research Work entitled Eldeuly Individual Fall Detection Using Deep learning in fulfilment of the requirement for the award of the Degree of Bachelor/Master of Technology in Software Engineering and submitted to the Department of Software Engineering, Delhi Technological University, Delhi is an authentic record of my/our own, carried out during a period from Jan to May 25 under the supervision of Dr. Abhilasha Sharma

The matter presented in this report/thesis has not been submitted by us/me for the award of any other degree of this or any other Institute/University. The work has been published/accepted/communicated in SCI/ SCI expanded/SSCI/Scopus indexed journal OR peer reviewed Scopus indexed conference with the following details:

Title of the Paper: Vision Based Elderly Human Fall Detection system using Movenet Pose Estimation and Hybrid CNN BiLSTM Architecture
Author names (in sequence as per research paper): PIYUSH AGARWAL, Dr. Abhilasha Sharma
Name of Conference/Journal: NCACSI, 2025
Conference Dates with venue (if applicable): 1st June 2025
Have you registered for the conference (Yes/No)?: Yes
Status of paper (Accepted/Published/Communicated): Accepted
Date of paper communication: 17th May 2025
Date of paper acceptance: 19th May 2025
Date of paper publication: Yet to Done

23/SWE/10
PIYUSH AGARWAL
Student(s) Roll No., Name and Signature

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I, further certify that the publication and indexing information given by the students is correct.

Abhilasha Sharma
Supervisor Name and Signature

Place: Delhi

Date: 26/05/25

NOTE: PLEASE ENCLOSE RESEARCH PAPER ACCEPTANCE/ PUBLICATION/COMMUNICATION PROOF ALONG WITH SCOPUS INDEXING PROOF (Conference Website OR Science Direct in case of Journal Publication).