Volume 9, Issue 5, May – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY232

Domain-Adaptive and Context-Aware Fall Detection Based on Coarse-Fine Network Learning

[1]Dr. G. Indumathi (Assistant Professor)
Department of Computer Science and Engineering,
SRM Institute of Science and Technology,
Ramapuram, Chennai, India

[2]A. Dinesh Kumar Reddy
Department of Computer Science and Engineering,
SRM Institute of Science and Technology,
Ramapuram, Chennai, India

[3]Anuvind Udayan Akral
Department of Computer Science and Engineering,
SRM Institute of Science and Technology
Ramapuram, Chennai, India

[4]M. Jaswanth
Department of Computer Science and Engineering,
SRM Institute of Science and Technology
Ramapuram, Chennai, India

Abstract:- Accurate fall detection among older adults is crucial for minimizing injuries and fatalities. However, existing fall detection systems face challenges due to the rarity and variability of falls, compounded by limitations in real-world datasets. To address this, a novel fall detection approach integrating domain adaptation and context-awareness within a Coarse-Fine Network Learning framework is proposed. The model combines high-level semantic understanding with low-level spatial details to achieve robust fall detection across diverse environments. Domain adaptation techniques like transfer learning and domain-specific fine-tuning are introduced to enhance model generalization and adaptability. Additionally, context-aware features, including environmental cues and behavioral patterns, reduce false alarms. Extensive experimentation on real-world datasets demonstrates the superior performance of the model, outperforming traditional approaches. The framework holds promise for deployment in healthcare settings, contributing to improved safety for older adults worldwide. The interpretability of the model's predictions enhances its usability in practical applications.

Keywords:- *ADL :Activities Daily Living, CNN :Convolution Neural Network.*

I. INTRODUCTION

Human falls pose a serious threat to the health of the elderly, the disabled, and those recovering from injuries.In recent years, there has been a lot of interest in the topic of human fall detection due to the need to support people who are in danger. The objective is to develop a system that can independently determine if a person has fallen and requires assistance. Systems use a variety of data sources to achieve this, including video data from cameras, ambient sensors like sound sensors, and sensor data from wearable technology. Advances in computer vision and video comprehension have led to an increased interest in vision-based systems for detecting falls in humans.

Among the senior population's primary causes of hospitalization and loss of independence are falls. Quick intervention could lessen the consequences of an incident involving falls. In order to facilitate the detection and triggering of a fall alarm, automatic fall detection systems have been developed in recent years. These systems mostly rely on wearables or smartphones with integrated inertial sensors and location capabilities. A supervised system for fall detection was developed using accelerometer signal processing in previous published methods using a dataset of simulated falls and activities of daily living (ADLs), which are regarded as non-falls, collected from young volunteers.Compared to the volume of daily activities, falls are exceedingly infrequent events. Additionally, the annotation process required to create a real-world falls dataset consumes a significant amount of time and resources, making such datasets highly rare. Nevertheless, a few study teams have been successful in gathering information from actual falls.

Falls are a common global health concern, especially for the older. 33% of grown-ups north of 65 fall in any event one time each year . worldwide. When a fall happens, there is a higher risk of damage and death if one is not physically strong and fit. Furthermore, a positive correlation can be shown between the waiting time for rescue and the death rate. The mortality risk rises to 50% among senior people who fall and lie on the ground for longer than 2.5 hours, or a "long lie." As a result, creating a trustworthy fall detection (FD) system is crucial for first aid applications.

More than thirty percent of the elderly have a fall at least once a year, making falls the second most common cause of accidental injuries worldwide behind traffic accidents. 60% of older people who fall suffer brain injuries, and 90% result in hip bone fractures. Furthermore, after a fall, almost half of the elderly stay on the ground for an extended period of time (long-lying), which increases the risk of pressure sores, dehydration, hypothermia, pneumonia, and even death. As a result, a fall detection system that can promptly alert nursing staff to senior patients who have fallen is essential. The fall detection

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY232

system can reduce the amount of damage from the fall and expedite the arrival of medical assistance.

Because life expectancy is rising and the birth rate is decreasing, population aging is becoming a global issue. The World Health Organization (WHO) projects that the number of elderly persons (60 years of age or over) will rise from 900 million in 2015 to around 20 billion by 2050, or 22% of the world's population. But as people age, their functional abilities inevitably decline. This includes physical, sensory, and cognitive impairments, all of which raise the risk of falling.An estimated 28–35% of seniors 65 years of age or more are said to fall every year. With age comes an increased danger of falling. According to reports, the percentage of senior citizens who fall every year and get moderate to serious injuries—such as bruising, hip fractures, or head trauma—has increased to 32-42%. In addition, falls exacerbate financial strains, psychological anguish, and even lower the quality of life for caregivers. Early intervention in the event of an elderly person's fall could lessen the devastating consequences.

Several studies in the sector have shown promising outcomes when employing manually created video features that are then combined with conventional classifiers. Simultaneously, the focus of generic video interpretation research has switched to deep learning techniques. It should come as no surprise that deep learning techniques have become more popular in vision-based human fall detection systems.

The United Nations projected that 13% of the world's population would be 60 years old or more seasoned, making the improvement of capabletechnologies to help and support the elderly all the more important. Falls pose a serious risk to the health of older folks, as they not only injure the elderly physically but even young people living alone. According to data from the World Health Organization1, 37.3 million falls are Sufficiently serious to require clinical consideration for recuperation,, and falls cause around 684,000 deaths annually.There are various approaches to fall detection, including employing video surveillance systems, wearable technology, and Wi-Fi signals. Compared to wearable technology, video-based systems are more practically implemented because they don't place as much physical strain on users or necessitate intricate operating procedures.To produce accurate findings, most video-based fall detection methods rely on action recognition models that have been proven to work.

Fall events usually result in a number of expensive aftereffects, such as:

- Causing significant harm to senior citizens, such as hip, arm, wrist, and ankle fractures.
- Those who are taking specific medications may find their circumstances worse if they sustain head traumas.

In addition, if an elderly person sustains head trauma from a fall, they should visit the hospital as soon as possible to check for brain damage.

- Making a lot of people afraid of falling and reducing their level of activity. They consequently get weaker and have a greater likelihood of experiencing the same scenario once over.

It is more important than ever to build a fall detection system because of the frightening consequences that falling can cause. Furthermore, one of the most important variables in assessing the severity of a fall in an elderly person is how long they stay on the floor following the incident. Early fall detection can lessen the negative effects of the incident by enabling older individuals to obtain timely assistance from caregivers. This encourages us to work on this project further.

II. LITERATURE SURVEY

- Human action recognition and posture prediction are critical components of computer vision, focused on identifying actions and predicting body postures in videos to enhance machine perception and enable intelligent interaction. Recent advancements in deep learning technologies have significantly propelled these fields forward, allowing for the development of advanced algorithms for action recognition and posture prediction.

This paper presents a comprehensive review of recent developments in these areas, including datasets, feature representation methods, and the application of advanced deep learning algorithms for action recognition and posture prediction.

These techniques demonstrate exceptional real-world performance with high accuracy and are capable of scaling to various dataset sizes and complexities while maintaining manageable control overhead for practical deployment. However, successful implementation requires specialized knowledge in computer vision and deep learning. Challenges related to scalability may limit their applicability in large-scale deployment scenarios, necessitating optimization for broader use cases.

- This paper explores skeleton-based human action recognition, focusing on learning optimal skeletal representations within the context of action recognition. Many existing approaches rely on predefined sensor-captured skeletons, which can be ineffective. The methodology involves reconstructing 3D meshes from RGB videos and using a transformer-based algorithm (referred to as SGN) to extract informative skeletal representations, aiming to enhance action recognition accuracy. Experimental results on challenging datasets, specifically the SYSU and UTD-MHAD benchmarks, demonstrate the superiority of learned skeletal representations over sensor-captured skeletons. The technology leverages computer vision techniques for 3D mesh reconstruction and the SGN algorithm for efficient skeletal representation extraction. Despite the efficiency gained through reduced computational complexity and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY232

simplicity, challenges remain in real-time implementation due to complexity and scalability issues observed with previous solutions. Continued innovation is required to address these challenges and facilitate broader practical deployment of this approach. The comparison with sensor-captured skeletons on these benchmark datasets highlights the effectiveness and potential impact of the proposed skeletal representation approach in advancing human action recognition tasks.

- In this research, Graph Convolutional Networks (GCN) are explored for modeling human body skeletons as spatial and temporal graphs, benefiting skeleton-based action recognition. Existing GCN-based methods face challenges in integrating graph-structured skeleton representation with other modalities, potentially limiting scalability and performance. Additionally, pose information, rich in discriminative clues for action recognition, is underexplored in current methods. To address these issues, a pose-guided GCN (PG-GCN) framework is proposed. This multi-modal approach incorporates a multi-stream network to extract robust features from pose and skeleton data, using a dynamic attention module for early-stage fusion. By leveraging a trainable graph to aggregate features from pose and skeleton streams, PG-GCN achieves enhanced feature representation, leading to state-of-the-art performance on NTU RGB+D 60 and NTU RGB+D 120 datasets. The method introduces novel algorithms for feature fusion, including pose-guided attention mechanisms to capture joint correlations and improve model capability. Overall, PG-GCN demonstrates significant advancements in skeleton-based action recognition through innovative algorithms and technology integration.
- Human activity acknowledgment (HAR) innovation is an essential region in human-PC collaboration, focusing on stable real-world applications. Our HAR system excels in identifying detailed actions across varied scenarios, as evidenced by experiments conducted on the UTD-MHAD dataset. Comparisons with previous research show our system achieves an impressive 91% average accuracy. We also explored using HAR for crime detection, surpassing benchmark datasets in accuracy.

Despite these successes, our system faces challenges. It can be time-consuming, and its opportunistic nature limits real-time implementation. Additionally, while efficient, the system's uncontrollable aspects pose challenges in operational predictability. To optimize performance, we emphasize the importance of detailed training data tailored to specific applications. Moreover, our experiments highlight the significance of utilizing angle data, which aids in recognizing consistent action patterns amidst data variations. This research underscores the critical role of training data and data type selection in developing robust HAR systems for practical deployment.

- Multimodal human action recognition using depth sensors is crucial for applications like healthcare monitoring, smart buildings, transportation, and security surveillance. However, challenges arise from sub-actions sharing, complicating action recognition. This paper proposes a segmental architecture that leverages sub-action relations, heterogeneous data combination, and Class-Security Protected Cooperative Portrayal (CPPCR) for improved recognition. The segmental architecture models long-range temporal structures to distinguish similar actions with shared sub-actions. Depth motion and skeleton features are extracted and fused, and CPPCR addresses sub-action sharing by enhancing within-class local consistency in Collaborative Representation (CR) coding. Experimentation on four datasets validates the method's effectiveness. Challenges remain, including limitations in configuration changes, potential computation burdens, and application performance issues that need further exploration for real-world scenarios. The approach emphasizes low deployment cost, simplicity, and speed, making it accessible and impactful for various applications.
- This paper presents a human motion prediction framework aimed at enhancing safety and effectiveness in human-robot collaboration (HRC). The framework focuses on predicting human arm motion trajectories during a reaching task, enabling proactive robot behavior and uncertainty characterization. By combining partial trajectory classification and human motion regression, the framework can recognize human actions and predict trajectories before completion. The offline phase involves training a relapse model with enhanced hyperparameters and a combination procedure to consolidate expectation calculations. In the web-based stage,, multi-step Gaussian process regression and representative trajectory are used for trajectory prediction based on partial motion classification. Results demonstrate significant performance improvement, with the framework achieving a well-balanced tradeoff among parameters. Despite its complexity in installation and maintenance, the approach efficiently reduces human effort and boosts performance, overcoming critical design challenges associated with proactive robot behavior in HRC.
- The GL-LSTM+Diff model proposed in this work enhances human action recognition using RGB-D camera-captured skeleton joints. Unlike traditional methods that treat all skeletal joints equally, this model introduces innovative components to address spatial and temporal variations in joint contributions. A Global Spatial Attention (GSA) model assigns varying weights to different joints, improving spatial precision. Additionally, an Accumulative Learning Curve (ALC) model highlights frames critical for decision-making by adjusting temporal weights. By integrating GSA and ALC into the LSTM framework, the GL-LSTM model recognizes actions with improved accuracy.

Furthermore, a preprocessing method called Diff enhances feature dynamics to extract distinguishable features in deep learning. Rigorous experiments on NTU RGB+D and SBU datasets demonstrate superior

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY232

performance compared to state-of-the-art methods, showcasing the effectiveness of GL-LSTM+Diff in action recognition. The proposed model offers enhanced spatial and temporal processing, outperforming classic methods like STA-LSTM while minimizing complexity and training overheads. The inclusion of Diff cues further enhances feature recognition and dynamic capabilities, validating the model's effectiveness across diverse action recognition tasks.

- This paper proposes a clever way to deal with human activity acknowledgment in view of skeleton information, focusing on improving representation through a data reorganizing strategy that captures worldwide and nearby design information of skeleton joints. The method utilizes data mirroring to enhance joint relationships and employs an end-to-end multi-dimensional CNN network (SRNet) to leverage spatial and temporal information effectively for robust feature extraction. Experimental evaluations on diverse action recognition datasets like NTU RGB+D, PKU-MMD, SYSU, UT-Kinect, and HDM05 demonstrate the superiority of this approach over existing methods. However, challenges include potential issues with big payloads, tedious message updating, and increased capital and operating expenditures despite gains in real-time feasibility and hardware resource consumption. The SRNet algorithm incorporates specialized convolutional operations tailored to the multi-dimensional structure of skeleton data to optimize feature extraction and improve action recognition accuracy.
- The proposed approach introduces a Composite Latent Structure (CLS) model for 3D human action representation and recognition. This model represents human actions as hierarchical graphs, with sequential atomic actions represented by composite latent states composed of both semantic and geometric attributes. discriminative EM-like calculation is utilized to get familiar with these composite dormant designsfrom skeleton sequences, enabling accurate action recognition and inference of latent temporal structures. The method is evaluated on testing 3D activity datasets including MSR 3D Activity Dataset, Multiview 3D Occasion Dataset, and UTKinect Activity 3D Dataset,, demonstrating its effectiveness and advantages. The algorithm leverages hierarchical graph representations and discriminative learning to model complex human actions with composite attributes, providing insights into latent structures for improved recognition performance.
- This research introduces a method for human action recognition that addresses challenges in traditional approaches, such as low accuracy and adaptability. The proposed method leverages deep neural networks and a novel parameter instatement strategy based on the multi-facet max out network initiation capability to alleviate hardships in training, such as gradient issues and overfitting. The approach involves detecting and tracking human actions, encoding temporal and spatial features using restricted Boltzmann machines (RBMs), and integrating these features into a global representation for action videos. Trained SVM classifiers are then used for action recognition. Experimental results demonstrate high accuracy and adaptability of the proposed method, showing promise for complex scene recognition and opening new avenues for feature extraction in human action analysis. However, challenges include the effectiveness of existing solutions and the limited configurability post-initialization.

III. PROPOSED MODEL

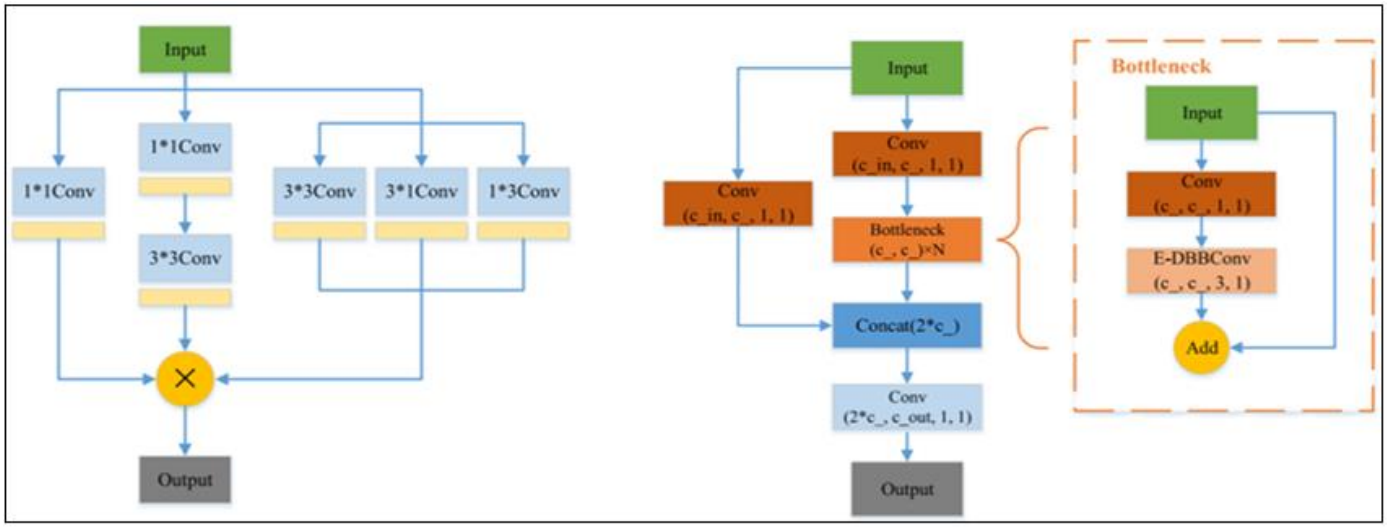

Fig 1 Proposed Model Partitions Our Dataset into Specific Proportions Showing CNN Layers

Figure 1. we partition our dataset into preparing set and test set by a specific proportion, and the number of preparing tests and test tests of each class subsequent to examining of the dataset. The proposed contains convolution layers, a nonlinear enactment capability, pooling layers, straightening, and fully associated layers. The convolution layers remove spatial highlights from the crude information with various channels sliding over the information. The

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY232

summation of the component wise increase of a channel and open field of the information is determined as the layer yield, as displayed in Fig. 1. Each channel is prepared utilizing back propagation and can address various parts of the information.Nonlinear enactment capabilities, for example, the corrected direct unit (ReLU), sigmoid (S), and exaggerated digression (tanh) capabilities, are applied to the results of the convolution and fully associated layers to work on the exactness of the classifier. The ReLU capability, communicated by Condition ReLU (x) = {0 x< 0 x ⩾ 0 Pooling layers decrease the elements of the information by applying specific standards inside the channel. Max pooling, the most widely recognized strategy, recovers the most extreme worth inside the channel as yield. After the convolution and pooling layers, the separated two-layered highlights are straightened into onedimensional portrayals to be input into the fully associated layers. A fully associated layer figures the result vector utilizing Condition y =σ(wx + b) The spine network creates the last highlight maps for RGB, profundity, and middle of the road modalities. We utilize the extension highlight misfortune to Compele the weighted amount of distances between the middle of the road methodology highlight guide and those of RGB and profundity. The weighted aggregate utilizes the weighting coefficients got from the beforehand referenced IDM module. This guarantees that when the RGB methodology affects the moderate methodology, the scaffold include misfortune places more prominent accentuation on the distance between the RGB also, middle modalities. Likewise, the equivalent standard applies to the profundity methodology.

➢ *Dataset Description:*

The datasets utilized for simulation purposes consist of raw RGB and Depth images resized from 640x480 to 320x240, captured from a single uncalibrated Kinect sensor positioned at approximately 2.4m above ground level. In total, the datasets comprise 21,499 images. Among the total of 22,636 images, 16,794 images are allocated for training, 3,299 for validation, and 2,543 for testing. These images are distributed across five distinct rooms, each containing eight different viewpoints. The dataset features five individuals: two males aged 32 and 50 and three females aged 19, 28, and 40. Each individual performs activities corresponding to five distinct classes of poses: standing, sitting, lying, bowing, and crawling. Each image contains only one individual, with some images labeled as empty or "other." For training and validation, images of a 32-year-old male and a 28-year-old female are utilized, totaling 16,794 and 3,299 images, respectively. The validation set includes images of the 32-year-old male from the training set but captured in a different room. The test set comprises images of three individuals: two females aged 19 and 40 and a male aged 50, all captured in a room distinct from the training and validation sets. The 22,636 images are sequenced without repetition, and each set includes horizontally flipped images to augment the dataset**.**

➢ *Image Preprocessing:*

Preprocessing procedures are a significant piece of the profound growing experience. Via carefully preprocessing information, we can work on the execution of our prepared model and make it more vigorous to varieties in the information. Additionally, it can essentially affect the presentation of the prepared model. There are an assortment of preprocessing procedures that can be utilized in profound learning, contingent upon the particular dataset and job that needs to be done. Some normal preprocessing methods that we utilized are as the followings:

- *Standardization:*

In the event that the info pictures are standardized, the model will combine quicker and more exact. At the point when the info pictures are not standardized, the shared loads of the organization have unique alignments for various elements, which can drive, the hour of cost capability to combine, taking more time and in less capably way. Normalizing the information makes the expense capability a lot simpler to prepare.

- *Flat Flip:*

It is a kind of change that is utilized in information expansion procedures to expand the dataset utilized in profound discovering that makes the pictures flip horizontally from left to right. It makes a reflected picture of the unique picture along the upward pivot.

Fig 2 An Optical Stream is Applied to the Subject in a Scene. (a) Subject Turning. (b) Subject Getting up after a Fall Occasion

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY232

Figure 2. Both depict a person in an indoor setting, possibly an office or a laboratory. Computers and desks are visible in the background. The person is wearing dark clothing and appears to be captured mid-motion.

➢ *Model Learning &Prediction:*

The preparation information was parted into preparing, approving, and testing segments. Scarcely any Ages were chosen for every cycle. A sum of 80% of the information was picked for preparing, while the leftover 20% was held for testing and approval. The learning rate for the last layers boundaries is 1e-03 and the learning rate for different layers boundaries is set to 1e-04. The cross-entropy misfortune and Adam enhancer are utilized with bunch size of 128 Pooling Layer Pooling Layer Once the element maps are identified, geography of information has previously protected in highlight maps. Subsequently, area data turns out to be less significant. Pooling activity is applied to lessen the goal of component maps and accomplish spatial invariant. Each element map in the pooling layer is relating to explicit component map in past layer. In this way, the quantity of component maps in current layer is equivalent to past layer. Completely Associated Layer Completely Associated Layer Dropout Layer Fu ly-associated layer usua ly comes after a few internal item layers, pooling layers and consolidating layer. Every hub (neuron) in fu ly-associated layer associates a l hubs in past layer. The result of hub is figured by adding al the loads duplicating hubs in past layer and going through actuation capability. Dropout Layer Dropout is a regularization procedure that zeros out the initiation upsides of haphazardly picked neurons during preparing. This imperative powers the organization to learn more hearty elements as opposed to depending on the prescient capacity of a sma l subset of neurons in the organization. Tompson et al. stretched out this plan to convolutional networks with Spatial Dropout, which exits whole element maps as opposed to individual neurons. Bunch Standardization Layer Group Standardization Layer Clump Standardization is one more regularization procedure that standardizes the arrangement of enactments in a layer. Standardization works by deducting the clump mean from every initiation and isolating by the bunch standard deviation. This standardization strategy, alongside normalization, is a standard procedure in the preprocessing of pixel values.

➢ *Classification Capabilities:*

As assessment measurements, we chose the precision and F1 scores. Exactness is the most essential measurement for a order task, while the F1 score joins recall and accuracy for a more thorough and natural assessment of an order model. Exactness, accuracy, recall, and F1 score can be determined by the fo lowing conditions: 1. Responsiveness & Explicitness Awareness and Particularity The responsiveness is comparable The True Positive Rate (TPR), also known as sensitivity, indicates the proportion of accurately recognized fall events out of the total number of actual fall events. Conversely, specificity refers to the rate of accurately identified non-fall events, which can be calculated as 1 minus the False Positive Rate (FPR). Sensitivity, synonymous with responsiveness and TPR, represents the ratio of correctly identified fall events (True Positives) to the sum of True Positives and False Negatives. Specificity, analogous to TNR and True Negative Rate (TNR), indicates the proportion of accurately identified non-fall events (True Negatives) to the sum of True Negatives and False Positives. These metrics are particularly relevant for datasets with inherent class imbalance, such as those related to fall detection, where positive (fall) cases are relatively infrequent compared to negative instances. Another method for evaluating model performance is through the Receiver Operating Characteristic (ROC) curve, a standard approach for illustrating a model's performance across various discrimination thresholds or hyperparameter values. Given the imbalance in the dataset, a precision-recall curve is utilized instead of the traditional ROC curve to effectively compare different model structures. Accuracy is calculated as the sum of True Positives and True Negatives divided by the total number of instances. In evaluating fall prediction models, the threshold for classifying a prediction as positive (indicating a fall) can be adjusted to accommodate specific application requirements. The values of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) instances are crucial for assessing model performance, with TP representing correctly predicted fall events, FP denoting incorrectly predicted fall events, TN indicating correct predictions of no fall events, and FN representing missed fall predictions. occasions happen. Bogus Negative (FN),the model predicts no fall occasions to happen, yet fall occasions happen.

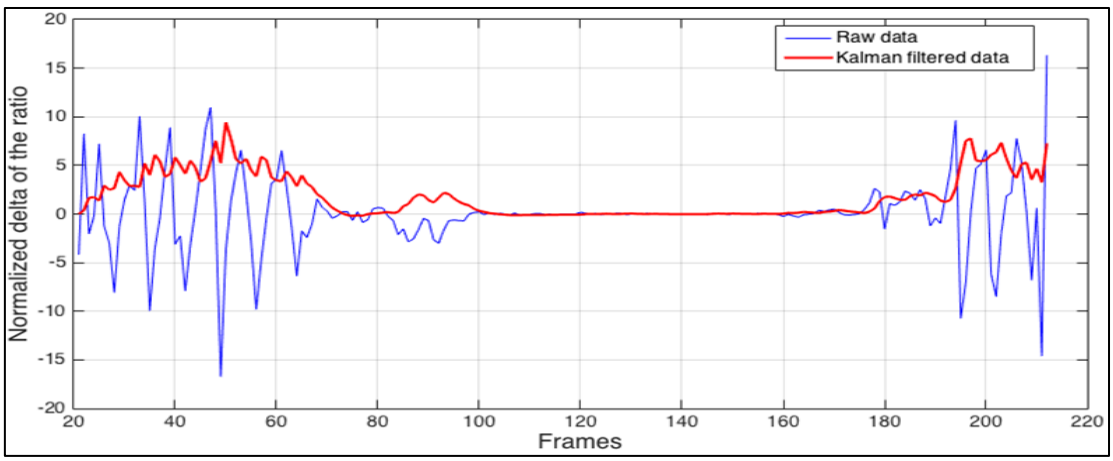

Fig 3 The Graph Shows Model Learning and Prediction of Variables

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY232

Figure 3.The timelines depicted in Figures 3 illustrate the subject's entrance into the scene at the outset. As the timeline unfolds, we observe the subject traversing the room, transitioning from a regular walking gait to a stable state of falling as the descent commences. This transitional phase is evident in both examples. Similarly, upon recovery from the fall, another transition occurs as the subject returns to an upright position. The intermediary phase, marked by consistent values, signifies the stable state of falling. In the final phase of both instances, the subject stands upright again and exits the scene. It's noteworthy that falls occurring at different angles relative to the camera exhibit distinctive characteristic values for the variables under observation.

IV. EXPERIMENTAL RESULT AND DISCUSSION

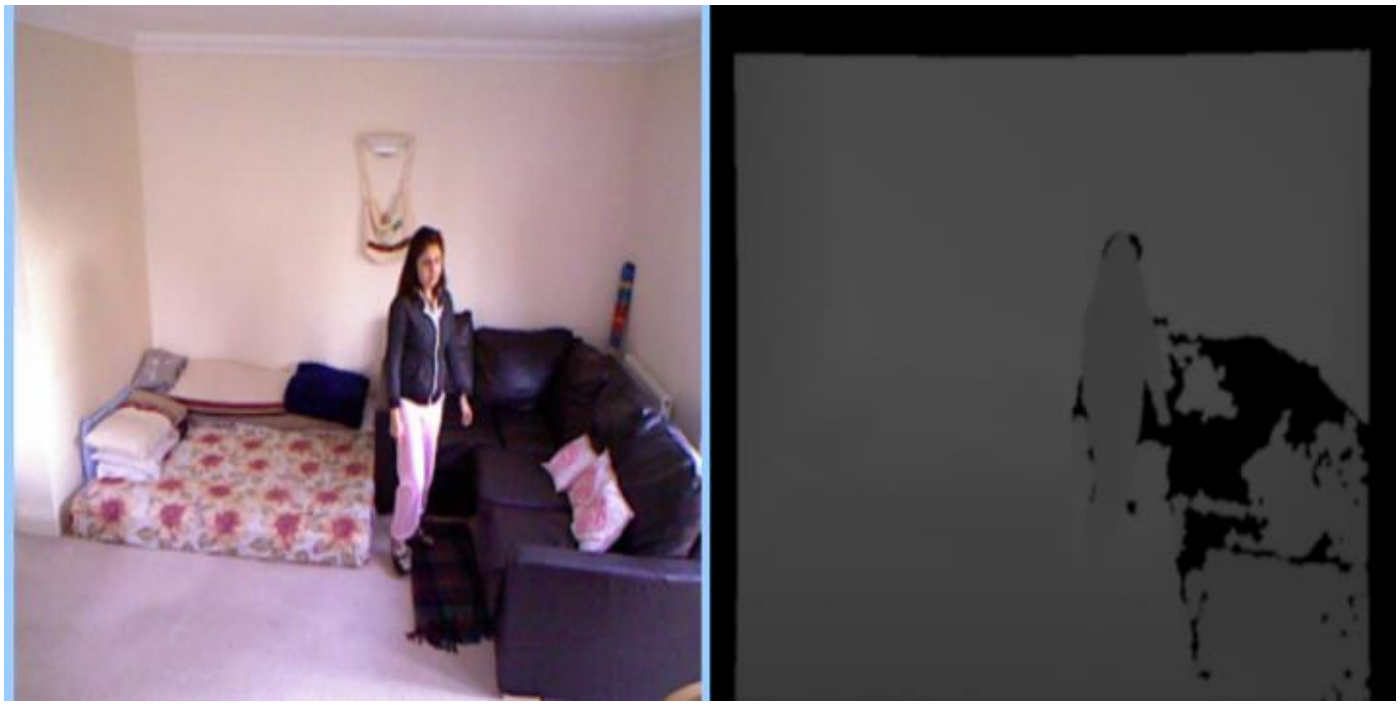

Fig 4 The Experimental Result Shows the Required Output with RGB Schemes

Figure 4. The image is split into two sections. On the left, there is a photograph of a room with furniture including a sofa, an armchair, and a bed with visible bedding.

The simulation utilizes raw RGB and Depth images, originally sized at 640x480, which are resized to 320x240. These images are captured by a single uncalibrated Kinect sensor fixed approximately 2.4 meters from the ground. The dataset comprises 21,499 images in total. Among the 22,636 images available, 16,794 are designated for training, 3,299 for validation, and 2,543 for testing. These images portray activities in five distinct rooms, each offering eight different viewing angles. The participants encompass five individuals, including two males matured 32 and 50, and three females matured 19, 28, and 40. The exercises performed by the members span five pose categories: standing, sitting, lying, bending, and crawling. Images categorized as 'other' represent empty scenes. For training, a dataset of 16,794 images is employed, featuring a male aged 32 and a female aged 28. Validation comprises 3,299 images, including a male aged 32 The dataset was meticulously crafted for the purpose of advancing research in vision-based indoor fall detection. Its creation stemmed from the recognition of a need within the field to accurately identify specific body poses indicative of falling incidents. By capturing a diverse range of poses in a controlled indoor environment, the dataset aims to provide researchers with a comprehensive set of images for training and testing fall detection algorithms. Each image in the dataset portrays individuals of varying ages and genders, captured in distinct rooms not present in the training or validation sets. Through augmentation techniques such as horizontal flipping, the dataset's size is bolstered to enhance its robustness and applicability. Each image is meticulously labeled with the corresponding pose classification, enabling researchers to evaluate the performance of their detection models across different scenarios. Overall, the dataset serves as a valuable resource for advancing the state-of-the-art in indoor fall detection through vision-based approaches.

V. CONCLUSION

This paper addresses the critical issue of fall risk detection, emphasizing the factors contributing to fall risk and the resulting healthcare costs. Falls pose significant dangers, particularly for the elderly, often leading to physical injuries and mental distress. Implementing preventive measures is essential to mitigate these risks, especially amid shortages of healthcare workers and the rising financial strain on healthcare systems.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24MAY232

The experimental findings demonstrate flawless performance in fall detection within controlled environments. The proposed method effectively minimizes false alarms and enhances fall detection precision, albeit requiring increased investment in cost and time.

This fall detection system is well-suited for deployment in expansive settings dedicated to elderly care, such as nursing hospitals and health centers, where robots can navigate efficiently. Its potential application promises enhanced safety and timely response to falls, ultimately improving the quality of care provided to vulnerable populations.

Continued advancements in fall detection technology will be pivotal in addressing the challenges posed by falls and ensuring prompt intervention, ultimately contributing to the well-being and independence of individuals at risk.

REFERENCES

[1]. Kaiqiang Huang,Susan Mckeever,Luis Mira les-PechuÃ¡n Generalized Zero-Shot Learning for Action Recognition Fusing Text and Image GANs IEEE Access, 2024

[2]. Junuk Cha,Muhammad Saqlain,Donguk Kim,Seungeun Lee,Seongyeong Lee,Seungryul Baek Learning 3D Skeletal Representation From Transformer for Action Recognition IEEE Access, 2022

[3]. Yun Han,Sheng-Luen Chung,Qiang Xiao,Wei You Lin,Shun-Feng Su Global Spatio-Temporal Attention for Action Recognition Based on 3D Human Skeleton Data IEEE Access, 2020

[4]. Nan Ma,Zhixuan Wu,Yiu-ming Cheung,Yuchen Guo,Yue Gao,Jiahong Li,Beijyan Jiang A Survey of Human Action Recognition and Posture Prediction Tsinghua Science and Technology, 2021

[5]. Jaeyeong Ryu,Ashok Kumar Patil,Bharatesh Chakravarthi,Adithya Balasubramanyam,Soungsi l Park,Youngho Chai Angular Features-Based Human Action Recognition System for a Real Application With Subtle Unit Actions IEEE Access, 2022

[6]. Chengwu Liang,Deyin Liu,Lin Qi,Ling Guan Multi-Modal Human Action Recognition With Sub-Action Exploiting and Class-Privacy Preserved Co laborative Representation Learning IEEE Access, 2022

[7]. Qinghua Li,Zhao Zhang,Yue You,Yaqi Mu,Chao Feng Data Driven Models for Human Motion Prediction in Human-Robot Co laboration IEEE Access, 2020

[8]. Weizhi Nie,Wei Wang,Xiangdong Huang SRNet: Structured Relevance Feature Learning Network From Skeleton Data for Human Action Recognition IEEE Access, 2019

[9]. K. Huang, L. Mira les-PechuÃ¡n and S. Mckeever, Enhancing zero-shot action recognition in videos by combining GANs with text and images, Social Netw. Comput. Sci., vol. 4, no. 4, pp. 375, May 2023.

[10]. A. Salazar, L. Vergara and G. Safont, Generative adversarial networks and Markov random fields for oversampling very sma l training sets, Expert Syst. Appl., vol. 163, Jan. 2021.

[11]. H. Ding, Y. Ma, A. Deoras, Y. Wang and H. Wang, Zero-shot recommender systems, arXiv:2105.08318, 2021.

[12]. L. Wang, D. Q. Huynh and P. Koniusz, A comparative review of recent kinect-based action recognition algorithms, IEEE Trans. Image Process., vol. 29, pp. 15-28, 2020.

[13]. J. Wang, Y. Chen, S. Hao, X. Peng and L. Hu, Deep learning for sensor-based activity recognition: A survey, Pattern Recognit. Lett., vol. 119, pp. 1-3, Mar. 2019.

[14]. A. U lah, K. Muhammad, I. U. Haq and S. W. Baik, Action recognition using optimized deep autoencoder and CNN for survei lance data streams of non-stationary environments, Future Gener. Comput. Syst., vol. 96, pp. 386-397, Jul. 2019.

[15]. IEEE Trans. Cognit. Develop. Syst., vol. 14, no. 1, pp. 246-252, Mar. 2022. J. Munro and D. Damen, Multi-modal domain adaptation for fine-grained action recognition, Proc. CVPR, pp. 122-132, Jun. 2020.

[16]. X. Qin, Y. Ge, J. Feng, D. Yang, F. Chen, S. Huang, et al., DTMMN: Deep transfer multi-metric network for RGB-D action recognition, Neurocomputing, vol. 406, pp. 127-134, Sep. 2020.

[17]. H. Wang, Z. Song, W. Li and P. Wang, A hybrid network for large-scale action recognition from RGB and depth modalities, Sensors, vol. 20, no. 11, pp. 3305, Jun. 2020.

[18]. A. K.-F. Lui, Y.-H. Chan and M.-F. Leung, Mode ling of pedestrian movements near an amenity in walkways of public buildings, Proc. 8th Int. Conf. Control Autom. Robot. (ICCAR), pp. 394-400, Apr. 2022.

[19]. W. Cao, Z. Zhang, C. Liu, R. Li, Q. Jiao, Z. Yu, et al., Unsupervised discriminative feature learning via finding a clustering-friendly embedding space, Pattern Recognit., vol. 129, Sep. 2022.

[20]. Y.Jiang, D. K. Han and H. Ko, Relay dueling network for visual tracking with broad field-of-view, IET Comput. Vis., vol. 13, no. 7, pp. 615-622, Oct. 2019.

[21]. Y. Jin, J. Hong, D. Han and H. Ko, CPNet: Cross-para lel network for efficient anomaly detection, Proc. 17th IEEE Int. Conf. Adv. Video Signal Based Survei l. (AVSS), pp. 1-8, Nov. 2021.