ActiSenSee: Discrete and Continuous Activity Detection and Transition Identification Using Smart Glasses

Mrittika Raychoudhury, [1] Haoxiang Yu,[2] James D. Kiper [1]
[1]Miami University Oxford, OH - 45056
[1]raychom@miamioh.edu, [1]james.kiper@miamioh.edu
[2]University of Texas at Austin, Austin, TX - 78713
hxyu@utexas.edu

Abstract

Many users especially older adults suffer from presbyopia which is a visual inability to focus on nearby objects. Potential solutions include using separate glasses for activities that require different focal lengths for near and distant vision. Usually, sedentary activities like sitting, reading, and browsing on the computer require near-vision glasses while ambulatory activities like walking, jogging, and driving require distant vision glasses. Users often forget to change glasses while switching from sedentary to ambulatory activities. Using near-vision glasses for ambulatory activities leads to impaired depth perception and contrast sensitivity, increasing the chances of falls. In order to detect user activities distinctly and to alert them to change glasses while changing between activities, we have proposed a novel activity detection system called ActiSenSee. ActiSenSee uses a pair of glasses augmented with inertial sensors to capture various discrete and continuous activities from 23 users. Applying various machine learning and deep learning-based algorithms shows that ActiSenSee can accurately identify the different activities and the key activity transition from sitting to standing in discrete and continuous settings.

Introduction

Developments in small sensing devices and short-range wireless communication protocols have significantly boosted many research areas and Human Activity Detection is a major one among them. Human activity detection has many applications including detecting [Li et al. 2009; Kerdjidj et al. 2020] and/or preventing fall incidents [Kosse et al. 2013] among older adults. The Centers for Disease Control and Prevention, USA, predicted that at the current rate, by 2030, there will be 7 deaths per hour caused by falls for older adults [Moreland, Kakara, and Henry 2020]. Falls are caused by various factors including loss of balance, medical conditions, as well as visual distortions caused by incorrect depth perceptions [Black and Wood 2005]. The last one is often caused by the use of improper glasses or improper use of those glasses. E.g., in the case of single-focal length (SFL) glasses, near-vision glasses that is used for reading are often not suitable for walking or taking stairs as they significantly distort the field of vision. Similarly, using distant vision glasses used for driving or jogging is not suitable for near-vision activities like reading. But the latter may not lead to falls if the near vision activities are performed by users while sitting. The same cannot be said for the former one as distant vision activities are usually associated with ambulatory (or mobility-related) activities.

Many users with multiple vision problems use multi-focal length (MFL) glasses which can be divided into bifocal, trifocal, and progressive lenses. Wearing MFL has increased in recent times as it has multiple benefits like using a single glasses for most of the daily activities requiring different focal lengths. However, researchers have identified several disadvantages of using MFL glasses [Haran et al. 2009] including the trip hazard caused by impaired depth perception and contrast sensitivity [Lord et al. 2002; Lord, Smith, and Menant 2010]. In fact, users of MFL are always advised to "avoid looking down while walking". To address this issue older adults are also advised to use separate single focal-length glasses for separate activities.

In this research, we plan to encourage users to use separate single-focal-length glasses for sedentary and ambulatory activities. To ensure users change into proper glasses while shifting from sedentary to ambulatory activities we aim to detect a sitting-to-standing activity transition through an inertial sensor-based system and generate an alarm to remind users. The IMU sensor is attached to a user's glasses (henceforth called "smart glasses") and continuously detects the user's activities and keeps looking for a desired activity transition event.

Wireless sensor-based activity recognition and fall detection are a mature research area [Kerdjidj et al. 2020; Kelly et al. 2002]. However, activity detection using smart glasses is novel. In this research, we proposed an ActiSenSee system and collected data from 23 users wearing a pair of smart glasses and performing two sedentary and two ambulatories (standing or walking-focused) activities in a stand-alone or discrete manner. One more activity was transitional - sitting to standing and vice versa. The sensor data is analyzed through a set of machine learning and deep learning algorithms to accurately detect the activities and the transitions. Finally, a set of continuous activities were performed by the users to ensure the ML model can also detect the continuous activities using their knowledge of the discrete activity patterns. Results show promising outcomes of the ActiSenSee system and the unique and novel contributions are listed below:

- Developed the first smart glasses-based activity detection system to alert users of single focal length glasses about activity transition.
- Achieved an accuracy of 89% using traditional ML algorithms and 95% using more data and computation-intensive deep learning methods for activity detection.
- Applied the cost-effective ML models for continuous activity detection and transition recognition on real data collected from 23 users and achieved the highest accuracy of 91%.

Related Work

In this section, we discuss related research in sensor-based activity detection and the effects of single and multi-focal glasses on users.

Wearable Sensor-based Human Activity Detection & Fall Prevention

Human activity detection is an interesting and important research area being developed since the 1990s [Foerster, Smeja, and Fahrenberg 1999]. Detecting human activities is significant for various applications. Researchers have adopted mainly two separate techniques [Lara and Labrador 2012] for activity recognition - (a) Using videographies or other external sensors and (b) using wearable sensors. Research works in the former category consider capturing videos and detecting user activities by analyzing those videos. Also, sensors like RFID tags are attached with daily used objects to detect user activities [Van Kasteren, Englebienne, and Kröse 2010; Tolstikov et al. 2011; Yang, Lee, and Choi 2011; Sarkar et al. 2011; Hong and Ohtsuki 2011]. Smartphone-based activity detection [Joseph et al. 2010] is also considered under this category. The latter category considers attaching various wearable sensing modules to the human body to keep track of the movement of various body parts. The most commonly used sensors are inertial measurement units (IMUs) like tri-axial accelerometers and gyroscopes. Moreover, the user's heart rate, blood pressure, location information, etc are also tracked to zero for proper activity determination by combining various sensor data. User locations in indoor and outdoor environments are tracked using GPS and WiFi signals. All relevant data from wearable sensor networks are integrated using smartphones and then forwarded to servers for further processing and storage. Also, the activity detection algorithms often operate in the server environment. One significant problem under the activity detection research is fall detection which is a major health concern for older adults in home or in assisted living [Ozcan et al. 2013; Zhu and Sheng 2011; Fortune et al. 2011]. There is extensive research coverage of fall detection literature in various survey articles [Kosse et al. 2013]. Fall prevention using wearable sensors has proposed several useful methods including a pressure-sensor augmented neoprene sock that senses pressure when the wearer stands up and alerts caregivers to assist the wearer to prevent fall risks.

Effects of Eye Glasses on Fall

Older adults often suffer from presbyopia which is a refractive error in eye lenses [Attebo, Ivers, and Mitchell 1999] causing difficulty in focusing on nearby objects [Donahue 1999]. To address this condition, older individuals use separate single-focal lens (SFL) glasses for different activities, such as reading glasses and driving glasses. This requires changing into proper glasses when a change in activity takes place. An alternative solution to mitigate the problem of changing glasses frequently is to use multi-focal lens glasses (MFL). MFL glasses can accommodate two distinct focal lenses (bifocal), three distinct focal lenses (trifocal), or continuously changing focal lenses (progressive). Users can use the lower part of the glasses for near-vision activities and the higher part for distant-vision activities without changing glasses. However, using MFL glasses have some problems for users in terms of impaired depth perception and contrast sensitivity [Lord et al. 2007, 2002]. Researchers have shown that older adults wearing MFLs find difficulty in step negotiation and proper foot placement while taking stairs [Johnson et al. 2008, 2007] which leads to around 8% increase in falls [Haran et al. 2010]. Brayton-Chung, et al. [Brayton-Chung, Tomashek, and Smith 2013] have established the differences in depth perception for users wearing SFL and MFL glasses.

Using Smart Glasses for Activity Detection

Researchers have worked for a long to develop wearable sensor-based activity detection mechanisms. However, using smart glasses for activity detection is novel and has not been investigated before. Although Google Glass [1] Vuzix glasses [2] have been introduced by companies, but there is no large-scale activity detection study using them except an eye blink detection with Google glasses [Ishimaru et al. 2014]. In our previous work, we have introduced this concept and have shown some basic experimental results with machine learning-based activity classification [Raychoudhury, Yu, and Kiper 2022] with a maximum accuracy of 87%. In this work, we conduct more extensive evaluations using ML and deep learning techniques and establish the usefulness of our proposed system in correctly recognizing user activities with accuracy up to 95%. Furthermore, we have also conducted continuous activity classification and transition detection with significantly promising results. Our approach is unobtrusive and efficient and can be adopted by older adults to remind them to change into proper glasses.

System Model of ActiSenSee

ActiSenSee proposed by us is an end-to-end system for user activity detection using smart glasses. Our target participants are users of single focal length eyeglasses who need to change glasses depending on the type of focal length required for the daily activity being performed. Users wear the sensor-augmented glasses and perform a set of activities listed in Table 1. Sensors embedded in the smart glasses (see Fig. 1) are a tri-axial accelerometer, a tri-axial gyroscope,

[1] https://www.google.com/glass/start

[2] https://www.vuzix.com/pages/smart-glasses

a magnetometer, and a pressure sensor. The IMU sensor is a MetaMotionC [3] wearable sensing entity which communicates with a smartphone app 'MetaBase' [4] from the same manufacturer using Bluetooth Low Energy wireless communication protocol. We chose MetaMotionC as the sensor incorporates a hardware-based Kalman filter to reduce noise and distortion in sensor data. The sensors collect raw data while the users perform the aforementioned activities and send those to the smartphone through an app. The sensor data is then uploaded to a server for cleaning and preprocessing. Processed data is then subject to a number of machine learning and deep learning-based classification algorithms and a suitable model is chosen for discrete activity recognition. The chosen model is applied to sensor data collected for a set of continuous activities and it can successfully recognize individual activities. The transition between sitting and standing is considered especially important for users and an alarm is generated to remind them to change into glasses with an activity-specific focal length.

Data Collection for Discrete & Continuous Activities

One major contribution of this research was the painstaking data collection process. We have collected data for 23 individuals who are healthy males and females aged 18 or older. The participants must have smartphones (although we provided the experimental smartphone) or are open to using smartphones and assistive technologies. They must be able to walk without any external support and must not have any uncorrected visual issues. Users who cannot satisfy the above requirements are excluded from this study. Participants were recruited from the students and staff of Miami University in Oxford, Ohio and this research is approved by the Miami University IRB protocol ID#02038r.

Potential subjects were invited to participate in the study via email, social networks, and flyers. Participation was completely voluntary. Declining to participate did not involve any penalty or loss of benefits to which they were otherwise entitled.

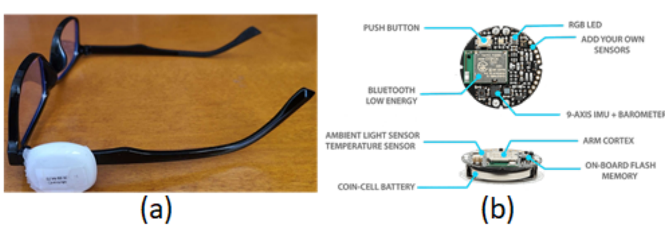


Figure 1: (a) Glasses augmented with IMU Sensors (b) Internal Structure of the IMU Sensor

Discrete Activity Data Collection

As already stated before, we have chosen a set of five distinct activities shown in Table 1 and asked the users to perform those while wearing the smart glasses. We used a mid-size room with a 15' x 12' containing a table and a chair.

Table 1: Class Labels of the 5 Distinct Activities (A1-A5)

	Activities	Labels
1	Sitting and Reading a Book	A1
2	Sitting and Working at a Computer	A2
3	Standing up from Sitting	A3
4	Walking	A4
5	Picking up Items from the Floor	A5
	Number of Activity Classes	**5**

The discrete activity dataset contains time-series data generated by a triaxial accelerometer, a triaxial gyroscope, a triaxial magnetometer, and a pressure sensor (Barometer). In Figure 2, we have shown the accelerometer data patterns for the five activities A1-A5. We can see that the patterns for sedentary activities are different from those of ambulatory and transitional activities.

Continuous Activity Data Collection

Continuous data collection refers to a procedure during which a participant performs a series of pre-specified activities while wearing sensor-augmented glasses. In order to standardize the set of activities for possible replication of the experiments to validate our results by any future researchers, we have decided to select the Timed Up and Go (TUG) Test.

Timed Up and Go (TUG) Test: The 'timed up and go test' (TUG) is a simple and popularly used clinical performance-based measure among the elderly to test functional mobility and detect the probability of falls [Nguyen et al. 2015]. The test requires only a few minutes, is easy to administer, and requires minimal equipment like a chair with a solid seat without any armrest and a flat back, a cellphone stopwatch, a distant marker (small red cones for our case), and proper walking shoes. The TUG test is used as a model of simple activities that consists of a continuous execution of four common activities shown in Fig. 3. To match our discrete activities, we have considered mainly the following four distinct TUG activities - Sitting, standing up, walking, and sitting down. The inertial sensor was placed firmly on the right temple of the glasses and data was collected from 23 healthy participants. The complete TUG test was performed 5 times with 10 seconds gap in between. However, the time to complete the entire TUG test is not measured by us as it is not a concern for our activity recognition research. The continuous data collection process was video recorded (with the participant's permission) for labeling purposes.

Processing and Analyzing TUG Test Data: At first, we labeled the video of the Continuous Activity through software called Lossless Cut[5]. We then matched the sensor data collected through the Metawear sensor with the video. Fig. 4 has the graphical visualizations of the TUG dataset of a person. We can observe how the signal values in the X, Y, and Z axes vary for a time period of 1 min and 20 seconds. For better visualizations we have filled in colors for each activity timeline, the green color area is for the sitting activity (A1),

[3] https://mbientlab.com/metamotionc/

[4] https://mbientlab.com/tutorials/MetaBaseApp.html

[5] https://github.com/mifi/lossless-cut

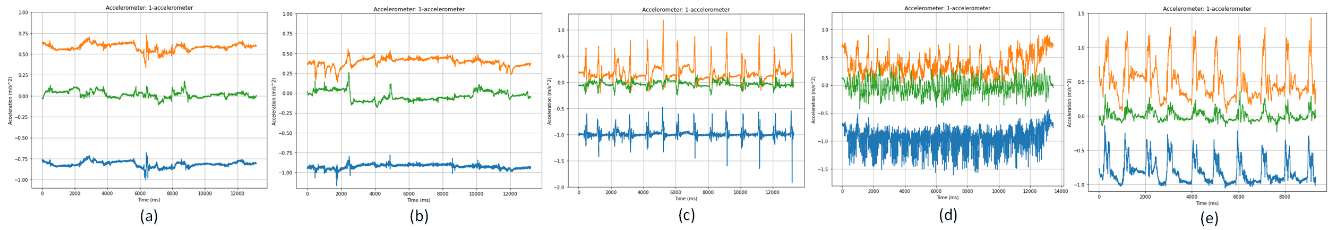


Figure 2: Accelerometer Sensor Data Pattern for activities (a) Sitting and Reading a Book, (b) Sitting and Working at a Computer, (c) Standing up from Sitting, (d) Walking, and (e) Picking up items from the Floor

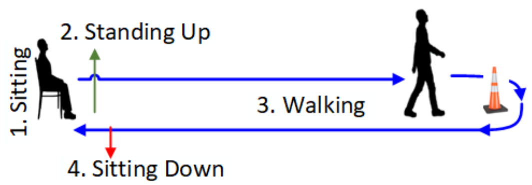


Figure 3: The TUG path and the major four activities performed during a TUG. These transitions are: 1) sitting, 2) sit-to-stand 3) walking, and 4) stand-to-sit

the plum for the sit-to-stand activity (A3), the light cyan for the walk and turn activity (A4) and pink for the stand to sit activity (A3). We have used the same activity labels as in the discrete data streams. Since sit-to-stand and stand-to-sit transitions are considered together as discrete activity A3, therefore, we have used the same label for both transitions in the continuous activity.

Figure 4: Analysing TUG video files of two participants. The continuous activities (color-coded)are (1) sit, (2) sit-to-stand, (3) walk, and (4) stand-to-sit. This cycle is repeated five times for each participant.

Experiment with Discrete Activities

In this section, we describe the different evaluation methods we have applied to the discrete activity time-series data and discuss the performance results with a thorough analysis. First, we describe the machine learning-based classification algorithms followed by the deep learning algorithms applied to the discrete activity data streams. Our machine learning algorithms are executed on a workstation computer having Intel(R) Xeon(R) W-2133 CPU @ 3.60GHz processor with 12 cores, 128 GB RAM, and a 64-bit Windows operating system. We used Python for data processing. The deep learning algorithms have been run on the GPU of Google Colab.

Discrete Activity Detection using Machine Learning

Data Segmentation and Feature Generation: After the raw data was collected via the different sensors, an important part of the pre-processing step was segmentation for feature extraction and training. We followed the sliding window technique, where we conducted data segmentation with 50% overlap of consecutive time windows [Lau and David 2010]. As the selection of the window size plays an important factor in classification performance, we have conducted experiments by varying window sizes from 1 - 5 seconds. As generating the right features plays an important role in machine learning, for each window we have handcrafted 22 different features like abstract values, means, and variances of accelerometer and gyroscope readings across the X-, Y-, and Z-axes. To select only the most relevant features, we applied dimensionality reduction using the *Maximum Relevance Minimum Redundancy* (MRMR) algorithm [Zhao, Anand, and Wang 2019] and selected a subset of 18 features having minimum correlation among each other.

Algorithms and Experiment: We performed 10-fold cross-validation on the training datasets (five datasets for five window sizes) with a random 80:20 split between the training and validation datasets. In order to increase the efficiency of the individual features on the training dataset we have applied a StandardScaler where the mean of the observed value is 0 and the standard deviation is 1. We used five different machine learning algorithms for discrete activity recognition - K Nearest Neighbor (KNN), Random Forest (RF), Support Vector Machines (SVM), Decision Tree (CART), and Naive Bayes (NB). We conducted hyperparameter tuning to find the best parameters for each of the five different classification algorithms and optimized our results. Once the best parameters were found, we retrained our model and carried out the final evaluation on the testing dataset to achieve the final results as discussed in the following section.

We ran the experiments on the five datasets for 10 times each and reported the mean corresponding to window sizes of 1 to 5 seconds.

Performance Results: Our results show that the best performance has been achieved for the 5-second window size for the Random Forest (RF) Classifier with 1000 number of trees as the hyper-parameter. For the RF model, the training accuracy was 88% and the validation accuracy was 89%

which showed that the model did not overfit. The ROC_AUC score with One versus Rest (OvR) of RF was also high with 98% for a 5-sec window.

For all activities (A1-A5) Random Forest has the best F1-Score with class A4 having the highest accuracy of 96%. We show the class-wise accuracy of RF with the 5-second window in Fig. 6a.

Comparison with the State-of-the-art: In our previous work in ActiviSee [Raychoudhury, Yu, and Kiper 2022] we used the same dataset and applied the same five machine-learning algorithms. However, the major difference was that we used TSFresh Package [Luqian and Yuyuan 2021] to compute 7924 features (both time and frequency domain) and applied Maximum Relevance Minimum Redundancy (MRMR) for dimensionality reduction to select 25, 50, 75, and 100 best features. We achieved the highest accuracy of 87% with SVM and 85% with RF. Since calculating 7924 features is resource intensive, in the current paper, we hand-crafted only a few simple time domain features and achieved an even higher accuracy of 89% (for the RF model) with proper parameter tuning. This is possibly due to the fact that the handcrafted feature selection method has a low computational cost and high discriminatory ability of time domain features. Since some researchers argue that shallow features are not enough to train a human activity recognition (HAR) model and struggles to recognize high-level or context-aware activities, we have also trained a few deep-learning models where the feature has been learned automatically from the raw data in the network. We achieved an accuracy of 95% with deep learning which is much higher than both our previous work and the current handcrafted feature-based discrete activity recognition model. In the following subsection, we discuss our deep-learning models in detail.

Discrete Activity Detection using Deep Learning

In Human Activity Recognition, generating handcrafted features is challenging as it depends on the researcher's domain knowledge and the performance of the machine learning techniques and is very much dependent on data representation[Ramasamy Ramamurthy and Roy 2018]. There are no universal methods for selecting the appropriate features. However, in deep learning, the raw data is directly fed to the model as input and the features are learned hierarchically from it by some non-linear transformations. Deep learning has been very popular in activity recognition, and the widely-used techniques include Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Long-short Term Memory (LSTM), Recurrent Neural Networks (RNN), etc. In this model, we have considered the X, Y, and Z axes of the accelerometer and gyroscope sensors (a total of 6 raw features) as the raw input to the deep learning model. Our target is to predict the class label i.e., the 'activity' that we want to predict. For the data cleaning process, we have dropped all the null values. We ran the training set of the 3 deep learning models for the discrete dataset for 23 participants for a minimum of 100 and a maximum of 400 epochs for a 1-second window with 80% overlap.

- Deep Neural Networks (DNN): The raw dataset was divided into two different train-validation splits - 80:20 and 70:30. Initially, we trained the dataset with a number of hidden layers from 1 to 5, varying the number of neurons per layer from 50 to 200. The batch size of the model was 100. We have used ReLU (Rectified Linear Unit) for the activation function, Adam Optimizer with a learning rate of 0.01, and Cross Entropy as the loss function. The DNN network is a good fit for the model as the gap between the training loss and validation loss is minimum. However, as shown in Table 2, we could not achieve good accuracy using DNN with either split even by varying the parameters.
- Convolution Neural Network (CNN): We have trained a CNN model with our dataset as CNN has several important advantages over other deep neural networks. CNN is noise resistant and is able to extract time-independent deep features. Here, the raw data from the accelerometer and gyroscope sensors were sampled in a fixed length of 1-second sliding window with 80% sample overlap (80 readings/window) between consecutive windows. For the experiment, we segmented a total of 1382240 accelerometer and gyroscope input data samples into input signal of batch size 100. The convolution blocks each consisting of 128 and 64 filters, computed the dilations between the cells with a kernel size of k = 3 * 3. After the 2 convolution layers, we have added a dropout layer of 0.5. After the dropout, the second convolution layer is followed by a max-pooling layer that calculates the maximum or largest element in the feature map. The pooling kernel size is 2. After the transformations and convolutions, the vector is flattened of complex features that have information about the original time series in a frequency of a wide range and time scale domains. The vector is then fed to a 2-layered fully connected neural network with 100 hidden neurons in the first layer and the Softmax activation function in the last layer. We considered 2 different splits of train-validation - 70:30 and 80:20, the results of which are shown in Table 2. We observed that the CNN model performed better than DNN.
- Long Short-Term Memory (LSTM): LSTM models are very popular for time-series forecasting. It learns a function that maps a sequence of past observations as input to an output observation. First, we split the dataset into 70-30 and 80-20 train and validation sets. As the LSTM expects a fixed-length sequence for training the model, we have used a sliding window technique where each window corresponds to 1 second of accelerometer and gyroscope readings. We have considered 80% of overlapping data and applied the Stacked LSTM to our labeled dataset directly maintaining the temporal sequence. We trained the model with a batch size of 128 and a learning rate of 0.01. Our Stacked LSTM architecture consisted of 3 hidden LSTM layers. The LSTM layer above provided the sequence output to the LSTM layer below. In order to prevent overfitting of the model and to improve its performance, we added a regularization method called the Dropout layer after every hidden layer. This probabilistically excluded the activation and weight updates dur-

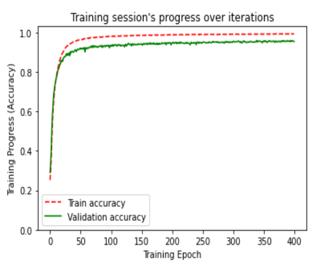
Table 2: Discrete Activity Detection using Deep Learning Models. Training and Validation Accuracies for LSTM, CNN, and DNN

DNN				CNN				LSTM			
70-30		80-20		70-30		80-20		70-30		80-20	
T	V	T	V	T	V	T	V	T	V	T	V
0.654	0.6495	0.658	0.653	0.962	0.837	0.968	0.856	0.998	0.945	0.997	0.952

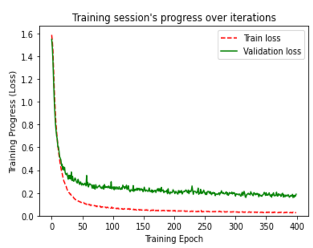
Table 3: Discrete Activity Detection using Deep Learning Models. Results for LSTM and CNN

3*ACTIVITY	CNN						LSTM					
	70-30			80-20			70-30			80-20		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
A1	0.83	0.84	0.83	0.87	0.84	0.86	0.95	0.94	0.94	0.94	0.96	0.95
A2	0.84	0.81	0.82	0.85	0.84	0.84	0.94	0.94	0.94	0.95	0.95	0.95
A3	0.85	0.83	0.84	0.87	0.84	0.86	0.95	0.94	0.95	0.95	0.95	0.95
A4	0.81	0.88	0.85	0.84	0.89	0.86	0.95	0.95	0.95	0.96	0.95	0.96
A5	0.85	0.83	0.84	0.85	0.87	0.86	0.94	0.95	0.94	0.96	0.95	0.95

ing the training phase and reduced the overfitting to 0.2. Without the Dropout, our model overfitted by more than 0.5. Two dense layers also known as the fully connected layer are added at the end of the neural network. The activation function used is ReLU and Adam Optimizer with a learning rate of 0.01. Our LSTM model achieved accuracies of 94% and 95% respectively for two different train-validation splits - 70:30 and 80:20 as shown in Table 2 with cross-entropy loss equal to 0.2. The change of training and validation accuracies and losses over iterations are shown in Fig. 5a and Fig. 5b.



(a) Change of Accuracies in LSTM



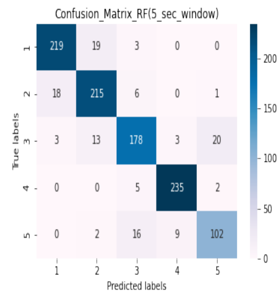
(b) Change of Loss in LSTM

Figure 5: Change of Train and Validation Accuracy and Loss Graphs for Best Deep Learning Model (LSTM) with 80:20 Split and 400 Epochs

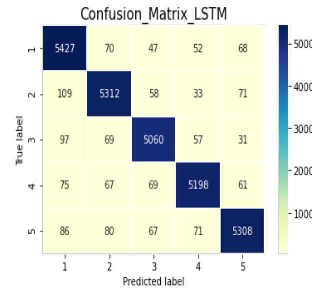
Summary of Results: In our research, activity A3 is the transitional activity from sitting to standing and is of utmost importance for alerting the user to change glasses. We can see that both for the discrete activity detection techniques using ML and DL models the F1 score of A3 is reasonably high. For, Random Forest ML model it is 86% and for the LSTM-based DL model, it is 95%.

Experiment with Continuous Activities

In order to test the robustness of our discrete activity detection models we conducted more experiments to successfully achieve continuous activity recognition with high accuracy using the cost-effective ML models discussed above.



(a) The Random Forest Confusion Matrix with 5-second window



(b) Confusion Matrix for LSTM with 80:20 Split and 400 Epochs

Figure 6: Confusion Matrices for Best ML model (RF) and Best DL (LSTM) Model.

We have experimented with the TUG test data of 7 participants and have discussed the results below.

Performance Results

For continuous activity recognition using the TUG Data, we have used 1-second sliding window-based data segmentation with 50% overlap and applied the Random Forest model which produced the best results for the discrete activity recognition. In TUG data, we focused on four primary activities - Sitting (A1), Standing up from Sitting (A3), Walking (A4), and Sitting from Standing (A3). We dropped discrete activities Sitting and Working at a Computer (A2) and Picking up Items from the floor (A5) from the training set as they are not relevant to the TUG activities. The RF model which was previously trained on the discrete activity dataset was used to predict the class labels of continuous activities from the labeled video data. We calculated the duration of each activity in seconds and tried to predict the label of the activity in that particular duration. We predicted the activities of 7 participants with maximum accuracy of 91%. The Precision, Recall, and F1-Scores of four sample participants have been shown in Table 5. Although the F1-Score was pretty good for A1 and A4, it suffered for A3. In order to increase the overall accuracy of the model we implemented a *Post-Process method* by adjusting model prediction based on common sense. Activity A3 should take place once between the sitting (A1) and walking (A4) activities. Thus, the Post-Process method added A3 in between A1 and A4, if it did not appear in the predicted label sequence (i.e, a person cannot start walking from a sitting position without standing up). Also, the A3 should not happen for more than 1 second. So, if there are multiple A3S occurring simultaneously, we converted the additional A3 to A1 or A4 depending on the previous activity.

In Fig. 7, the top one in red represents the raw predicted label sequence using the RF model on the TUG dataset; the middle one in green represents the label sequence after applying the Post-Process method; and the bottom figure represents the ground truth. We observe that the model incorrectly predicted some activity as A3, and believe that incorrect prediction is caused by the differences in the manner in which

Table 4: Mean and Standard Deviation of Accuracy for RF-predicted and Post-Processed Result of 7 participants for Continuous Activity Detection

2*Accuracy	Pre-Processed Result	Post-Processed Result
	0.73 + 0.11	0.78 + 0.12

Table 5: Precision, Recall and F1-Score of 4 Participants for Continuous Activity Detection

2*ACTIVITY	Participant1			Participant2			Participant3			Participant4		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
A1	1.00	0.95	0.97	0.81	0.84	0.82	0.89	0.89	0.89	1.00	0.64	0.78
A3	0.70	0.70	0.70	0.50	0.30	0.37	0.71	0.45	0.56	0.47	0.73	0.57
A4	0.92	0.95	0.94	0.93	0.98	0.95	0.92	0.98	0.95	0.94	1.00	0.97

different users perform a particular activity. However, those incorrect predictions have been effectively corrected by the post-process method using logical reasoning.

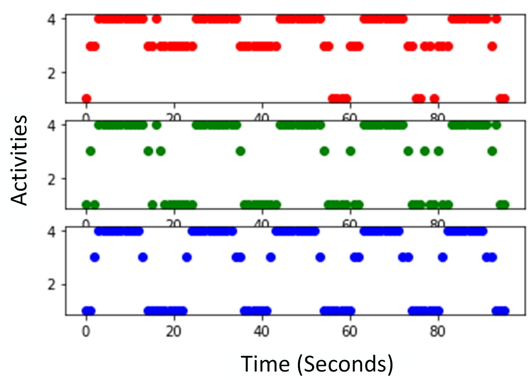


Figure 7: RF Model Predicted Label Sequences [Red]; Post-Processed Label Sequences [Green]; Ground Truth Label Sequences [Blue]

After applying the Post-Process method we could see from Table 6 that the accuracy did improve for some of the participants though not for all, like for participants 2 and 3 it did get better. From Table 4 we can see that the mean increased for 7 participants.

Discussion & Conclusion

Users with visual impairments who use multiple single focal length glasses for different daily activities which require different focal lengths (like sedentary and ambulatory activities) must remember to change from near to distant vision glasses as and when required in order to avoid trip hazards. Since users often forget to change glasses, it may lead to falls and injuries which are fairly common. In this research, we have collected sensor data from 23 healthy subjects while they perform a number of discrete and continuous activities wearing a pair of smart (sensor-augmented) glasses. The inertial sensor data streams while classified using certain machine learning and deep learning algorithms show promising results (89-95% accuracy) in distinctly recognizing the activities and the transitions between them. We have further applied one of the classification models to uniquely identify

Table 6: Change of Accuracy after Applying Post-processing Method

Participant 1		Participant 2		Participant 3	
Prediction Accuracy	Post-Processed Accuracy	Prediction Accuracy	Post-Processed Accuracy	Prediction Accuracy	Post-Processed Accuracy
0.911	0.911	0.652	0.863	0.773	0.892

continuous activity data streams and have achieved an accuracy of 91%. In the future, we plan to extend our research to include more participants and various other activities. Overall, the ActiSenSee system proves to be useful for users who use multiple singe-focal length glasses for different activities.

Acknowledgements

We earnestly thank Dr. Md Osman Gani (the University of Maryland Baltimore County) and Dr. Roger O. Smith (University of Wisconsin Milwaukee) for the motivation and concept development of this work. We also thank Miami University Centre for Assistive Technology (MUCAT) for providing hardware for our experiments.

References

Attebo, K.; Ivers, R. Q.; and Mitchell, P. 1999. Refractive errors in an older population: the Blue Mountains Eye Study. *Ophthalmology*, 106(6): 1066–1072.

Black, A.; and Wood, J. 2005. Vision and falls. *Clinical and Experimental Optometry*, 88(4): 212–222.

Brayton-Chung, A.; Tomashek, D.; and Smith, R. O. 2013. Fall Risk Assessment: Development of a Paradigm to Measure Multifocal Eyeglass Effects. *Physical & Occupational Therapy In Geriatrics*, 31(1): 47–60.

Donahue, S. 1999. Loss of accommodation and presbyopia. *Ophthalmology*, 11–21.

Foerster, F.; Smeja, M.; and Fahrenberg, J. 1999. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in human behavior*, 15(5): 571–583.

Fortune, E.; Tierney, M.; Scanaill, C. N.; Bourke, A.; Kennedy, N.; and Nelson, J. 2011. Activity level classification algorithm using SHIMMER™ wearable sensors for individuals with rheumatoid arthritis. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 3059–3062. IEEE.

Haran, M. J.; Cameron, I. D.; Ivers, R. Q.; Simpson, J. M.; Lee, B. B.; Tanzer, M.; Porwal, M.; Kwan, M. M.; Severino, C.; and Lord, S. R. 2010. Effect on falls of providing single lens distance vision glasses to multifocal glasses wearers: VISIBLE randomised controlled trial. *Bmj*, 340.

Haran, M. J.; Lord, S. R.; Cameron, I. D.; Ivers, R. Q.; Simpson, J. M.; Lee, B. B.; Porwal, M.; Kwan, M.; and Severino, C. 2009. Preventing falls in older multifocal glasses wearers by providing single-lens distance glasses: the protocol for the VISIBLE randomised controlled trial. *BMC geriatrics*, 9(1): 1–8.

Hong, J.; and Ohtsuki, T. 2011. A state classification method based on space-time signal processing using SVM for wireless monitoring systems. In *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, 2229–2233. IEEE.

Ishimaru, S.; Kunze, K.; Kise, K.; Weppner, J.; Dengel, A.; Lukowicz, P.; and Bulling, A. 2014. In the blink of an eye: combining head motion and eye blink frequency for activity recognition with google glass. In *Proceedings of the 5th augmented human international conference*, 1–4.

Johnson, L.; Buckley, J. G.; Harley, C.; and Elliott, D. B. 2008. Use of single-vision eyeglasses improves stepping precision and safety when elderly habitual multifocal wearers negotiate a raised surface. *Journal of the American Geriatrics Society*, 56(1): 178–180.

Johnson, L.; Buckley, J. G.; Scally, A. J.; and Elliott, D. B. 2007. Multifocal spectacles increase variability in toe clearance and risk of tripping in the elderly. *Investigative ophthalmology & visual science*, 48(4): 1466–1471.

Joseph, C.; Kokulakumaran, S.; Srijeyanthan, K.; Thusyanthan, A.; Gunasekara, C.; and Gamage, C. 2010. A framework for whole-body gesture recognition from video feeds. In *2010 5th international conference on industrial and information systems*, 430–435. IEEE.

Kelly, K. E.; Phillips, C. L.; Cain, K. C.; Polissar, N. L.; and Kelly, P. B. 2002. Evaluation of a nonintrusive monitor to reduce falls in nursing home patients. *Journal of the American Medical Directors Association*, 3(6): 377–382.

Kerdjidj, O.; Ramzan, N.; Ghanem, K.; Amira, A.; and Chouireb, F. 2020. Fall detection and human activity classification using wearable sensors and compressed sensing. *Journal of Ambient Intelligence and Humanized Computing*, 11(1): 349–361.

Kosse, N.; Brands, K.; Bauer, J.; Hortobagyi, T.; and Lamoth, C. 2013. Sensor technologies aiming at fall prevention in institutionalized old adults: A synthesis of current knowledge. *International Journal of Medical Informatics*, 82(9): 743–752.

Lara, O. D.; and Labrador, M. A. 2012. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, 15(3): 1192–1209.

Lau, S. L.; and David, K. 2010. Movement recognition using the accelerometer in smartphones. In *Future Network and Mobile Summit, 2010*, 1–9. IEEE.

Li, Q.; Stankovic, J. A.; Hanson, M. A.; Barth, A. T.; Lach, J.; and Zhou, G. 2009. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. In *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, 138–143. IEEE.

Lord, S. R.; Dayhew, J.; Sc, B. A.; and Howland, A. 2002. Multifocal glasses impair edge-contrast sensitivity and depth perception and increase the risk of falls in older people. *Journal of the American Geriatrics Society*, 50(11): 1760–1766.

Lord, S. R.; Sherrington, C.; Menz, H. B.; and Close, J. C. T. 2007. *Falls in Older People: Risk Factors and Strategies for Prevention*. Cambridge University Press, 2 edition.

Lord, S. R.; Smith, S. T.; and Menant, J. C. 2010. Vision and falls in older people: risk factors and intervention strategies. *Clinics in geriatric medicine*, 26(4): 569–581.

Luqian, S.; and Yuyuan, Z. 2021. Human Activity Recognition Using Time Series Pattern Recognition Model-Based on Tsfresh Features. In *2021 International Wireless Communications and Mobile Computing (IWCMC)*, 1035–1040. IEEE.

Moreland, B.; Kakara, R.; and Henry, A. 2020. Trends in nonfatal falls and fall-related injuries among adults aged $\geq$ 65 years—United States, 2012–2018. *Morbidity and Mortality Weekly Report*, 69(27): 875.

Nguyen, H. P.; Ayachi, F.; Lavigne-Pelletier, C.; Blamoutier, M.; Rahimi, F.; Boissy, P.; Jog, M.; and Duval, C. 2015. Auto detection and segmentation of physical activities during a Timed-Up-and-Go (TUG) task in healthy older adults using multiple inertial sensors. *Journal of neuroengineering and rehabilitation*, 12(1): 1–12.

Ozcan, K.; Mahabalagiri, A. K.; Casares, M.; and Velipasalar, S. 2013. Automatic fall detection and activity classification by a wearable embedded smart camera. *IEEE journal on emerging and selected topics in circuits and systems*, 3(2): 125–136.

Ramasamy Ramamurthy, S.; and Roy, N. 2018. Recent trends in machine learning for human activity recognition—A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4): e1254.

Raychoudhury, M.; Yu, H.; and Kiper, J. D. 2022. ActiviSee: Activity Transition Detection for Human Users through Wearable Sensor-augmented Glasses. In *Adjunct Proceedings of the 10th international workshop on human activity sensing corpus and applications (HASCA) in conjunction with ACM UbiComp*.

Sarkar, J.; Lee, Y.-K.; Lee, S.; et al. 2011. GPARS: A general-purpose activity recognition system. *Applied Intelligence*, 35(2): 242–259.

Tolstikov, A.; Hong, X.; Biswas, J.; Nugent, C.; Chen, L.; and Parente, G. 2011. Comparison of fusion methods based on dst and dbn in human activity recognition. *Journal of Control Theory and Applications*, 9(1): 18–27.

Van Kasteren, T.; Englebienne, G.; and Kröse, B. J. 2010. An activity monitoring system for elderly care using generative and discriminative models. *Personal and ubiquitous computing*, 14(6): 489–498.

Yang, J.; Lee, J.; and Choi, J. 2011. Activity recognition based on RFID object usage for smart mobile devices. *Journal of Computer Science and Technology*, 26(2): 239–246.

Zhao, Z.; Anand, R.; and Wang, M. 2019. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 442–452. IEEE.

Zhu, C.; and Sheng, W. 2011. Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(3): 569–573.