Machine Vision and Applications (2023) 34:124
https://doi.org/10.1007/s00138-023-01475-2

ORIGINAL PAPER

Vision-based approach to assess performance levels while eating

Muhammad Ahmed Raza[1] · Robert B. Fisher[1]

Received: 31 March 2023 / Revised: 22 September 2023 / Accepted: 23 September 2023 / Published online: 20 October 2023

Abstract
The elderly population is increasing at a rapid rate, and the need for effectively supporting independent living has become crucial. Wearable sensors can be helpful, but these are intrusive as they require adherence by the elderly. Thus, a semi-anonymous (no image records) vision-based non-intrusive monitoring system might potentially be the answer. As everyone has to eat, we introduce a first investigation into how eating behavior might be used as an indicator of performance changes. This study aims to provide a comprehensive model of the eating behavior of individuals. This includes creating a visual representation of the different actions involved in the eating process, in the form of a state diagram, as well as measuring the level of performance or decay over time during eating. Also, in studies that involve humans, getting a generalized model across numerous human subjects is challenging, as indicative features that parametrize decay/performance changes vary significantly from person to person. We present a two-step approach to get a generalized model using distinctive micro-movements, i.e., (1) get the best features across all subjects (all features are extracted from 3D poses of subjects) and (2) use an uncertainty-aware regression model to tackle the problem. Moreover, we also present an extended version of EatSense, a dataset that explores eating behavior and quality of motion assessment while eating.

Keywords EatSense · Motion assessment · Performance-level assessment

1 Introduction

On a global scale, the proportion of people aged 60 or over was just 8% in 1950, but this is projected to rise to 20% by 2050 [5]. The number of people growing older is increasing, whereas the increase in the number of caregivers is not proportionate. In a study carried out by Redwood et al. [41], it was reported that, in 2010, the caregivers ratio was more than 7 caregivers (including informal carers, such as family members) for every person in the high-risk age i.e., 80-plus. By 2050 the ratio of caregivers to seniors (i.e., seniors living in the community) will decrease to less than 3 to 1. With this growing burden, healthcare systems are under pressure and the situation of care homes is depressing as they have inadequate facilities [14, 38]. However, smart senior homes can be a potential solution that will not only help seniors to live independently more safely but also monitor their health status.

Seniors require constant monitoring and evaluation of their health and motor movements [24]. Unfortunately, periodic checkups and irregular motion analyses do not monitor the health status of an individual well enough. Reliable health profiling can only be done by constant monitoring with sufficient situational diversity. To bridge this gap, a variety of passive and active sensors have been proposed [24]. In this paper, we present a vision-based system that monitors a person while they eat and can assist in the early diagnosis of motor deterioration.

Why eating? Eating is one of the main, regular, and most important actions of one's daily life, so this is an opportunity for regular monitoring. We believe that monitoring the sub-actions of eating can provide evidence of major anomalies such as the presence or start of a neurological disorder or deterioration/decay of movement over time.

In this paper, we explore several research questions: What actions do people perform while they eat? Can we observe and distinguish gradual decay in motion over time while relying only on the camera as a sensor? Can we develop generalized models over all age groups for decay classifica-

✉ Muhammad Ahmed Raza
m.a.raza@ed.ac.uk

Robert B. Fisher
rbf@ed.ac.uk

[1] School of Informatics, The University of Edinburgh, Edinburgh EH8 9AB, UK

tion/regression as there might not be any consistent pattern to exploit across all subjects?

To answer these questions, firstly, we demonstrate through trunk stability and speed of movement tests that decay in performance is observable when weights of different levels are attached to the wrists of the subjects (Sect. 5). Secondly, we present a generalized model with strictly explainable features across various subjects in all age groups (Sect. 6).

For the results presented here, we propose an extension of EatSense [40], which is a human-centric, upper-body-focused dataset that supports the modeling of eating behavior as well as the investigation of changes in motion/motor decline (i.e., quality of motion assessment). Four levels of weights are put on the volunteers' wrists while they eat to simulate a change in mobility. The weights are not intended to be a model for aging, but only to demonstrate that minor changes in motion are detectable. The contributions of this paper are:

- The first computer vision-based quality of motion assessment quantitative approach solely based on the eating behavior of individual subjects.
- A state model for eating micro-movements[1] that represents the most common eating behavior among subjects of all ages (see Sect. 4).
- Address the most common problem of lack of generalizability when it comes to modeling human behavior (limited to the performance of eating assessment in our case). (see Sect. 6).
- Demonstrate that 4 weight classes simulate decay in the upper-body movements.
- Present the extension of the quality of motion assessment capability beyond EatSense by introducing a new abstraction level to the labels for each video (see Sect. 3)

2 Literature review

A brief review of past clinical and sensor-based techniques for decay assessment and behavior analysis is presented. Some publicly available benchmark datasets for motion quality assessment are also discussed.

2.1 Decay assessment tests

There have been many studies that list a set of tests in a clinical setting to observe decay in the functional motor movements [13, 15]. Alonso et al. [1] summarize clinical tests, such as 'timed up and go' and 'Functional Reach Test,' and computerized methods, such as 'Equitest' and 'Force Platforms' for assessing one's balance.

In a non-clinical setting, there also has been research that explores inertial measurement unit (IMU) or magnetometer-based motion tracking and assessment techniques. Filippeschi et al. [10] presented a survey where they compare IMU-based human motion tracking techniques with a focus on upper-body limbs which is potentially useful for motion assessment. Carnevale et al. [8] focused on shoulder kinematics assessment via wearable sensors after neurological trauma or musculoskeletal injuries. Recently, Meng et al. [27] presented an IMU-based upper limb motion assessment model and achieved good results.

Also in a non-clinical setting, there have been many vision-based healthcare results on (1) motion tracking, (2) fall detection [2], (3) anomaly in gait detection [51, 54], (4) exercises that help in the rehabilitation of people recovering from any disease that directly impacts their activity levels [4, 18, 42].

Nalci et al. [29] proposed a computer vision-based alternative test for functional balance that was compared with a BTrackS Balance Assessment Board (used in clinical assessments) to demonstrate the effectiveness of their proposed approach. Yang et al. [53] proposed a cost-effective and portable decision support system that used a single camera to track joint markers of upper-body limbs, perform data analytics for rehabilitation parameters calculation, and provide a robust classification suitable for home healthcare. In [22, 25, 30] the authors proposed a real-time risk assessment rapid upper-body limb assessment tool using cameras (depth or RGB) to detect anomalous postures in real-time and offline analysis.

Recently, Barlett et al. [3] proposed a vision-based balance assessment test while sitting. However, to the best of our knowledge, no vision-based study exists that explores decay/deterioration strictly based on the movement of upper-body limbs with the human pose.

2.2 Behavior analysis

Human behavior analysis is a broad term that deals with gesture recognition, facial expression analysis, and activity recognition. Onofri [33] suggests that activity recognition-based behavior analysis algorithms require knowledge that can be divided into two categories: contextual knowledge and prior knowledge. Contextual knowledge pertains to the context in which the action is taking place, such as the objects involved or the time and place. Prior knowledge is that the recognition system is aware of the past, such as event C frequently happens after event B, and the probability of event C happening after A is very low.

[1] Micro-movements, or sub-actions, refer to the individual and basic actions that are combined to form a single action. For instance, eating can be seen as a single action that involves several sub-actions, such as bringing the hand to the mouth.

Many studies have investigated human motion in sports games [7, 35, 49] and other applications [23, 26, 37]. Combining human body characteristics such as position, distance, speed, acceleration, motion type, and time is often used to quantify and evaluate behaviors. Oshita et al. [35] extracted the spatial, rotational, and temporal characteristics of the major poses of tennis trainees and compared their exercise patterns with experts.

In [55], to monitor a person's daily kitchen activities, Yordanova et al. presented a method for recognizing human behavior called Computational Causal Behavior Models (CCBM). This combined a symbolic representation of a person's behavior with probabilistic inference to analyze the person's actions, the type of meal they are preparing, and its potential health effects. Kyritsis et al. [21] introduced an algorithm that can automatically detect food intake cycles that occur during a meal using inertial signals from a smartwatch. They use five specific wrist micro-movements to model the chain of actions involved in the eating process 'pick food,' 'upward,' 'downward,' 'mouth,' and 'other movements.'

Previous research such as [32, 56] that utilize eating actions are mostly done for the sake of individual action understanding, i.e., to classify eating/drinking actions. On the other hand, in Tufano et al. [48] presents a systematic comparative analysis of 13 frameworks including deep learning and optical flow-based frameworks. The study focuses on detecting three specific eating behaviors, such as bites, chews, and swallows.

However, we are not aware of any previous studies analyzing eating behaviors and assessing the quality of motion based on those characteristics.

2.3 Public datasets for healthcare

Numerous openly accessible datasets explore certain aspects of healthcare. A few of them are discussed below.

Objectively Recognizing Eating Behavior and Associated Intake (OREBA) [45] is a dataset to offer extensive data collected from sensors during communal meals for researchers interested in the detection of intake gestures. OREBA includes various types of sensors, such as a 360-degree camera mounted at the front to capture video, as well as a sensor box that contains a gyroscope, an IMU, and an accelerometer attached to both hands. Other studies such as [21, 28, 46] also present small-scale datasets mainly focused on intake gestures, chews, and swallow behavioral characteristics.

Mobiserv-AIIA [17] was created to assess the intake of meals to prevent undernourishment or malnutrition. The collection includes recorded films that were made in a controlled laboratory setting using many cameras positioned at different angles. It entails employing a variety of tools while engaging in activities like eating and drinking for several meals (breakfast, lunch, and fast food) with using different tools to pick or scoop the food (spoon, fork or glass of water, etc.). The MSR-DailyActivity dataset [50] was created to simulate the day-to-day activities of a person sitting on a couch. It includes 320 examples of 16 daily activities such as 'play guitar' and 'eat.' RGB and a depth sensor were used to collect the MSR-DailyActivity dataset.

Sphere [36] was designed for motion quality assessment via gait analysis. Six participants were observed in this dataset, while they ascended a set of stairs. Init Gait DB [34] is a benchmark dataset for gait impairment research. The movement of limbs and body posture were changed to simulate eight various walking types. Several view angles were captured utilizing RGB cameras. The gait analysis-based walking dataset [31] replicates nine different walking gait patterns. This was recreated by attaching weights to the ankle or making one shoe with a thicker sole. This was captured using Microsoft Kinect where the participants walked on a treadmill with two flat mirrors behind them.

To the best of our knowledge, none of the existing datasets besides EatSense (discussed in the next section) provide the capability to assess the motion quality of humans with an emphasis on eating behaviors and a focus solely on the upper body joints.

3 EatSense

Aging has adverse effects on the musculoskeletal strength levels of all living beings, i.e., the older one gets, the motions of limbs slow down, postural control lessens, and hand-eye coordination gets tough. However, eating is an essential activity that everyone has to do regularly even in bad times. We presented EatSense, a novel dataset [40] that explores two areas in particular, i.e., sub-action recognition and quality of motion assessment. EatSense tries to address a few major research gaps, (1) sub-action recognition: The dataset has three levels of label abstraction and labels sub-actions with 16 classes where some of them only occur for less than a second, (2) sub-action temporal localization in videos that contains over a hundred subactions (on average) per video, (3) human-centered (hand gestures/posture based) eating behavior understanding, (4) decay in motor movement, i.e., small changes in upper-body movements, caused by attaching weights to the wrists of the subjects. However, previously, data were limited to only the binary classes 'weight' and 'no weight' (Y/N) at that time.

In this research, we present an extended version of EatSense[2] that simulates this decay in movement on a finer scale. Thus we expand our decay assessment classes by adding four different sizes of weights to the wrist, i.e., 0, 1 kg, 1.8 kg, and

[2] https://groups.inf.ed.ac.uk/vision/DATASETS/EATSENSE/.

2.4 kg. We also demonstrate the effectiveness of weights to simulate decay in Sect. 6.3.

3.1 EatSense collection and labeling

An RGB-Depth camera, Intel RealSense D415 was mounted on a wall at an oblique view angle in a dining/kitchen environment. The subjects were allowed to eat; however, they preferred without any external input from the recording team. The field of view had only one person at the dining table. EatSense contains 135 videos (53 for 0 kg, 25 for 1 kg, 33 for 1.8 kg, and 24 for 2.4 kg) with dense labels (all frames labeled without any stride). These videos are recorded at 15 frames per second (fps) with 640×480 resolution. Altogether, there are 705,919 labeled frames. Figure 1 shows the setting of the camera system in one of the dining room environments. It also shows one sample from the dataset both with and without wrist weights.

EatSense contains several labels for various levels of abstractions, i.e., (1) both 2D (extracted with HigherHRNet) and 3D (2D poses projected into 3D space using depth maps) for 8 upper body joint positions, (2) manually labeled 16 sub-actions for all frames in the videos, (3) binary labels based on if the subject is wearing a weight or not, i.e., 'Y'/'N'. The extension introduces a new level of abstraction, i.e., labels based on the weight a subject is wearing on their wrists, i.e., 0 kg, 1 kg, 1.8 kg, 2.4 kg.

Initially, we store both depth maps and RGB images. We employ Deep Privacy [16] to disguise the real face of the subjects in RGB videos to obscure their identity. The processed RGB, depth maps, and 3D skeletons are available to the general public for research.

3.2 EatSense properties

EatSense has many interesting properties that make it distinguishable from other existing datasets.

Dense Labels There are no unlabeled temporal patches in any of these videos, in contrast to the majority of large-scale datasets currently available. Additionally, a two-stage label quality control process enhances label consistency and reduces label errors.

Human-Centric Actions EatSense contains very consistent backgrounds and human posture-centric action examples, in contrast to other available datasets where background/environment can play a key role in differentiating between distinct actions.

Healthcare Analytics EatSense has a wide range of data that may be utilized to analyze human health. For instance, it has a layer of labels that can simulate (by the increase of weights) the gradual loss in a person's motor function over time. Continuously keeping an eye on the person's eating behavior and searching for signs of motor function decline may help save lives and identify the need for assistance before the situation gets worse.

3.3 EatSense feature extraction

For the purpose of exploration in the domain of health care, we propose and compute explainable hand-crafted features for EatSense and also compare them with deep features.

3.3.1 Hand-crafted features

The purpose of exploring hand-crafted feature-based techniques is to have an in-depth understanding of the individual subject's health. Deep features are convoluted and do not effectively help health professionals to understand the root cause of health problems faced by individuals. The proposed features are extracted over all individual frames.

These include instantaneous spatial features such as (1) relative distances of all joint locations concerning the chest, (2) relative joint locations in polar coordinates, (3) angles between shoulders and elbows, (4) product of all joints, (5) distance from the table of all joints. Also, temporal features such as (1) velocity, (2) acceleration, and (3) lags (past instantaneous joint position, i.e., if the current frame is captured at time t and we denote the joint position at t as $\mathbf{x}_t$, then the joint position in the previous frame taken at time $t - n$ denoted as $\mathbf{x}_{t-n}$ is the nth lag), (4) weighted sum of the last three lags. The mathematical formulation of each of these features is similar to that in [40].

Fig. 1 Left) the eight upper-body joints (1) nose (n), (2) chest, (3) right-shoulder, (4) right-elbow, (5) right-wrist, (6) left-shoulder, (7) left-elbow, and (8) left-wrist. Middle) subject is performing 'eat it' action without weights. Right) subject is performing 'eat it' action with weights

3.3.2 Deep features

For deep feature extraction for the videos in EatSense, a Spatial–Temporal Graph Convolutional Network (ST-GCN) [52] was used. In this approach, similar to the hand-crafted features, we exclusively utilize the 3D poses of the subjects. As previously discussed, HigherHRNet was used to estimate 2D poses from RGB data which were then projected into the 3D space with the help of depth maps, to estimate 3D joint location.

However, unlike the manual feature extraction, which operates on a frame-by-frame basis, we consider an entire action that extends across several frames to leverage both spatial and temporal characteristics to construct a graph. High-level feature maps are estimated by applying graph convolutions on the constructed graph.

4 Eating behavioral model

The EatSense dataset's sequences are densely labeled with 16 sub-actions of variable lengths to represent the eating behavior of individuals. Figure 2 presents a general state diagram showing the sequential relationships between the 16 sub-actions.

Upon examination, it becomes evident that the diagram allows much situational diversity, including a single-hand eating with or without a tool, two hands eating with or without a tool, and if the subject switches between either of these.

The eating behavior model illustrates that the actions 'eat it' and 'drink' consistently occur after the action 'move hand toward mouth' and are subsequently followed by the action 'move hand away from mouth.' Since the video recordings were acquired in an uncontrolled environment, the subjects were permitted to engage in conversations and use mobile phones, just as they would in their routine. Consequently, the state diagram demonstrates that nearly all actions can be followed by the activity labeled as 'other.'

5 Decay simulation

This section demonstrates the effectiveness of simulating decay in performance by adding different weights to the wrists of the subjects. For this purpose, experimentally proven tests such as the balance assessment and speed of motion tests are used. These tests are slightly modified

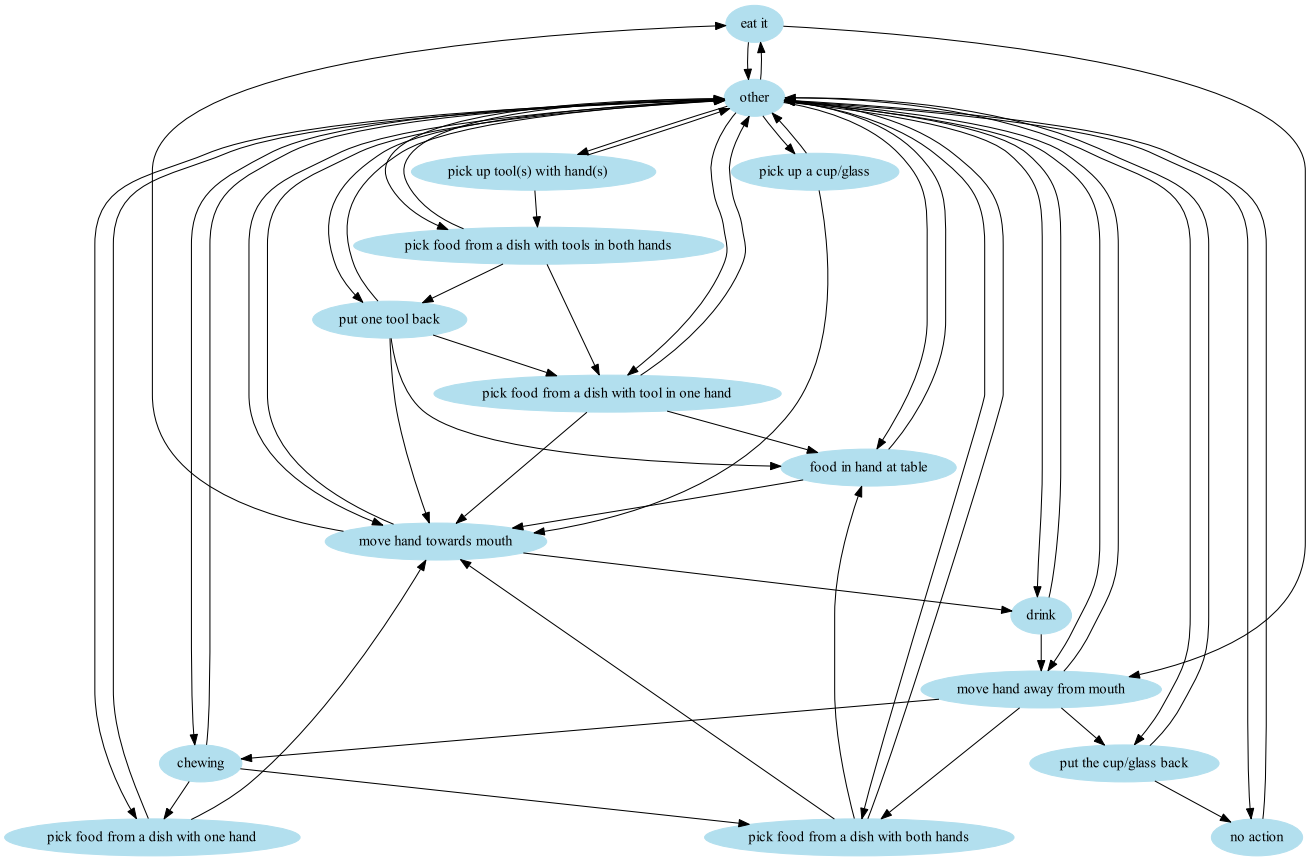

Fig. 2 State diagram of common eating behavior with 16 action classes

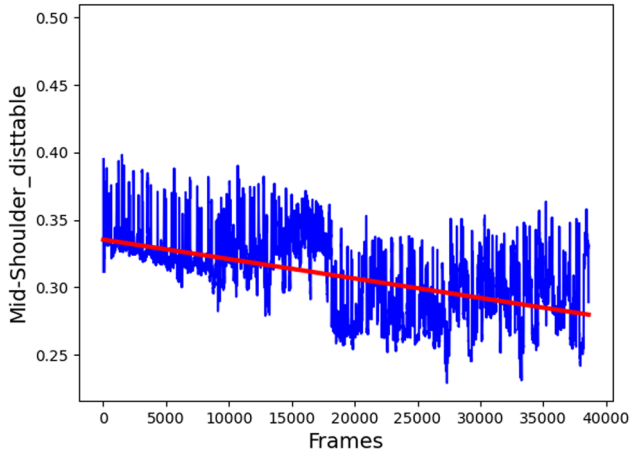

Fig. 3 Shown for demonstration of negative slopes only. This chart indicates 20% frames sampled randomly from each of the 4 weight cases of subject no. 4. These frames are then subsequently arranged in ascending order of their respective weights

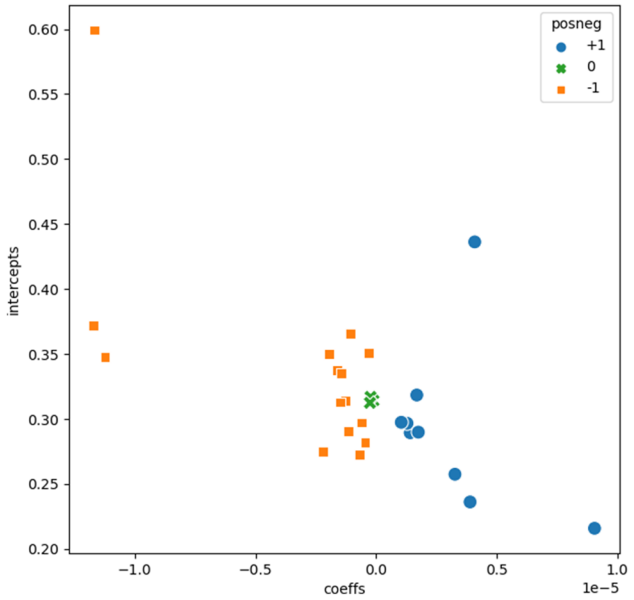

Fig. 4 Balance Assessment Test. +1 (blue) represents subjects with positive slopes, − 1 (orange) represents subjects with negative slopes, and 0 (green), which indicates a change in their trunk positions, i.e., the subjects started with an upright posture but over time as the weight is increased, their chest position changed. See the text for more discussion (color figure online)

according to the need of exploring decay in an eating scenario. These subtle changes along with the plots are explained in the sub-sections below.

5.1 Balance assessment test

The Balance Assessment Test [3, 20] also known as trunk stability or postural sway [6] test is defined as how well the subject maintains the center of mass of their body within its base support. In clinical trials, this is carried out while standing up; however, here the test is performed, while the person is seated for about 6–10 min for a full meal. Each of the subjects is recorded while wearing weights ranging from 0 to 2.4 kg in each individual video.

At every frame, using the 3D pose of the subject, we estimate the feature 'the distance of the chest with respect to the table' (discussed in Sect. 3.3) to detect sway in the subject's posture. As videos are recorded with participants wearing weights, we temporally stack the videos one after another in the increasing order of the weights. Two of the subjects were left-handed which were flipped around the y-axis for consistency.

Linear regression fits a line through the temporal data (videos stacked in the order of increasing weights). This is shown in Fig. 3 for demonstration purposes. The predicted line (shown in red) depicts a negative sloped line. The decrease in distance from the table while increasing weights is indicated by a negative slope. Hence, the negative slope in the experiment indicates the decay in performance as the weights are increased.

A negative slope indicates decay in the core/trunk position over time, and a positive slope should mean that the posture got better over time.A plot depicting the relationship between slope coefficients and intercepts is shown in (Fig. 4) where + 1 (blue) represents positive slopes, − 1 (orange) represents negative slopes; and 0 (green) represents no visible change in their trunk position. Here, visible change is measured and marked as either blue or orange if the coefficients are greater or less than $\pm 0.03 \times 10^{-5}$. The plot reveals that the majority of the subjects, specifically 15 out of 27, exhibit negative slopes. This indicates a weakened core as they were unable to maintain an upright position. On the other hand, a few subjects demonstrate a positive trend, which leads us to hypothesize that this occurs when they attempt to compensate for the weights by adjusting their balance.

5.2 Speed of motion test

The speed of motion test is based on how fast a subject performs a task at hand in their normal routine to monitor muscle degradation due to aging. The age-based decay in muscle functionality is known as sarcopenia [43, 44]. In this research, different levels of weights are used to simulate this decay in muscle strength over time and quantify it by monitoring the speeds of the motion of the upper body limbs.

Firstly, as the dataset contains multiple sub-actions, many of which include unpredictable orders of motion, only the 'move hand toward mouth' sub-action is analyzed, as it is the main micro-movement that involves motion against gravity. For this purpose, we estimate (by inter-frame position differences) the velocity of the dominant hand using the dis-

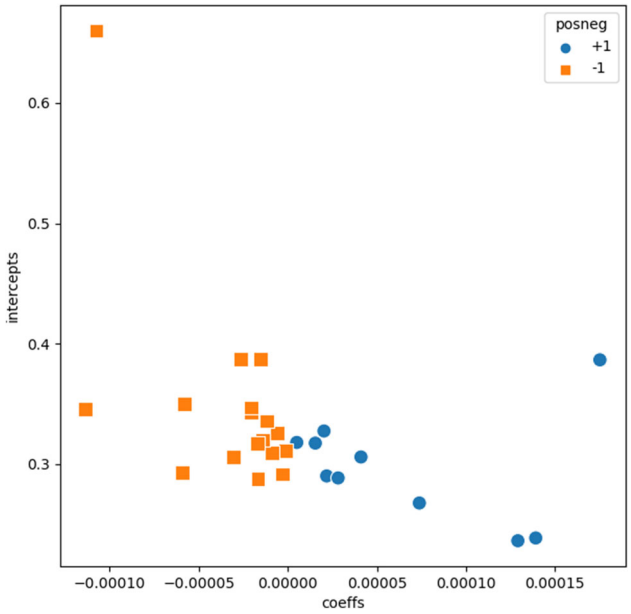

Fig. 5 Speed of Motion Test. 0 (blue) represents subjects with positive slopes, and 1 (orange) represents subjects with negative slopes, which indicates a decrease in hand speed as the weight is increased. See the text for more discussion (color figure online)

tance of wrist joint position relative to the chest (discussed in Sect. 3.3). Two of the subjects were left-handed which were flipped around the y-axis for consistency. Similar to the postural sway test, the wrist velocities are estimated in the increasing order of the weights. A line is fit through the speed versus weight curves for each subject using linear regression.

The slopes are expected to be negative to demonstrate that there is a decay in the upward movement speed. In Fig. 5, a scatter plot illustrating the relationship between slope coefficients and intercepts indicates that 17 out of 27 subjects exhibit a decline in their motion speeds across various weight classes. Conversely, the subjects who show either positive or neutral trends in the data are predominantly those who report having an active lifestyle.

6 Generalized regression

EatSense simulates decay by adding weights (i.e., 0 kg, 1 kg, 1.8 kg, and 2.4 kg) to the wrists of the subjects. These subjects belong to various ethnicities, genders, and age groups. Ideally, there would exist motion model with a common set of parameters to predict performance as weights are increased. However, it seems that people react to the weight increase differently; for example, some slouch more and some make a distinctly visible posture difference (dropped shoulders, etc.). Hence, finding a set of features and a model that parametrizes the performance change process without over-fitting on a subset of subjects is a problem. To model how performance changes with weight level, we divide our experiments into two sub-experiments, i.e., deep features-based and hand-crafted features-based regression.

6.1 Hand-crafted features-based regression

Both spatial and temporal features were extracted from joint locations. These are briefly discussed in Sect. 3.3, and their detailed mathematical formulation is given in [40]. The complete pipeline of the regression approach is shown in Fig. 6. The primary aim of delving into hand-crafted feature-based techniques is to gain a comprehensive understanding of an individual subject's health. By utilizing these techniques, researchers and health professionals can obtain detailed insights into various aspects of a person's well-being. On the other hand, deep features, although powerful in their ability to represent intricate patterns and relationships in data, have thus far not proven to be as conducive to providing interpretable explanations. Health professionals often seek clear and understandable insights into the factors influencing a subject's health, and in this regard, the complexity of deep features might present a challenge in meeting that need.

6.1.1 Feature selection

To select a common subset of features across all subjects, a forward sequential feature selector (FSFS) was used [39]

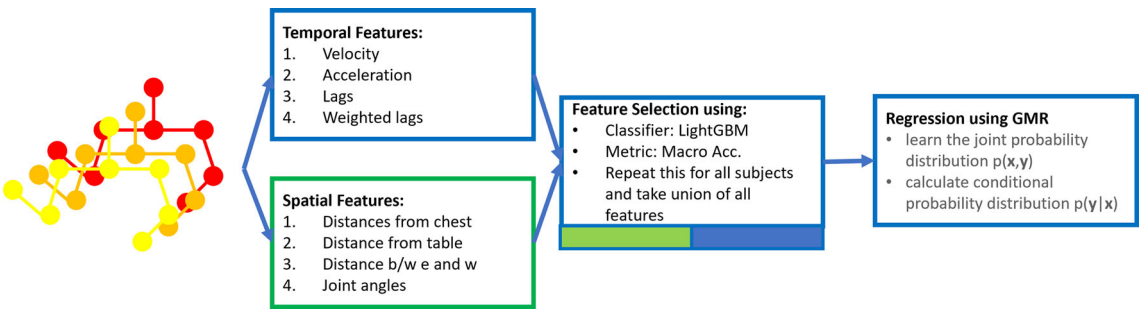

Fig. 6 The complete pipeline of the proposed regression approach

with LightGBM [19] as the classifier of the four classes of different weights in subsets of the dataset. Assume D represents the data comprising the subjects' joint locations relative to the chest and the rest of the features. A set f_i of the top-most contributing 12 features for each subject i, was selected based on maximum macro-accuracy.

Afterward, a union of f_i was taken to create a collection of 30 features. Finally, the forward sequential feature selector (FSFS) method was employed, using GMR as a regressor and mean-squared error as the loss function, to identify the top 8 most significant features (F) from this set of 30 (which were all used in the LightGBM, GMR, and MLP regression experiments in Sect. 6.3).

The process for feature selection across all subjects is shown in Eqs. 1 and 2, where $d_i \subset D$ is the subset that contains data for the ith subject only ($i = 1, \ldots, 27$). The subscripts C and R under FSFS show that the first FSFS used a classifier and the second used a regressor to shortlist the best set of features.

$$f_i = \text{FSFS}_C(d_i)_{i=1}^{27} \tag{1}$$

$$F = \text{FSFS}_R(\cup_{i=1}^{27} f_i) \tag{2}$$

The shortlisted 8 features in the order of their contribution are: (1) distance of the left-wrist from the table, (2) position at time t of the x-component of the left-wrist, (3) position at time $t-1$ of the y-component of the right-shoulder, (4) distance of table to the right-elbow, (5) position at time t of the y-component of the left-wrist, (6) distance of table to the left-shoulder, (7) velocity of x-component of the left-shoulder computed with window-size of ± 2, and (8) distance of table to the left-elbow.

The selected features contain both spatial (instantaneous distance from the table, position at time t, etc.) and temporal properties (position at time $t-1$, velocity, etc.). One noticeable trend is that most of the selected features are related to the left-arm. This highlights that the non-dominant arm plays a significant role for discriminating between different weights. This potentially indicates that with weights of different magnitudes, the movement of non-dominant arm appears to suffer from a more noticeable change then the right arm. This may be attributed to the fact that individuals typically employ their dominant arm for eating, as it is more accustomed to precise motor tasks and possesses greater strength.

6.1.2 Feature visualization

To illustrate how the data look like with 8 most contributing features, we project the 8-dimensional data to 2 dimensions using T-SNE. The data are visualized in Fig. 7. Although there are not four clearly separable groups for the four weights, there is somewhat of a clustering (especially for the red/no weight class) that suggests that some modeling is possible.

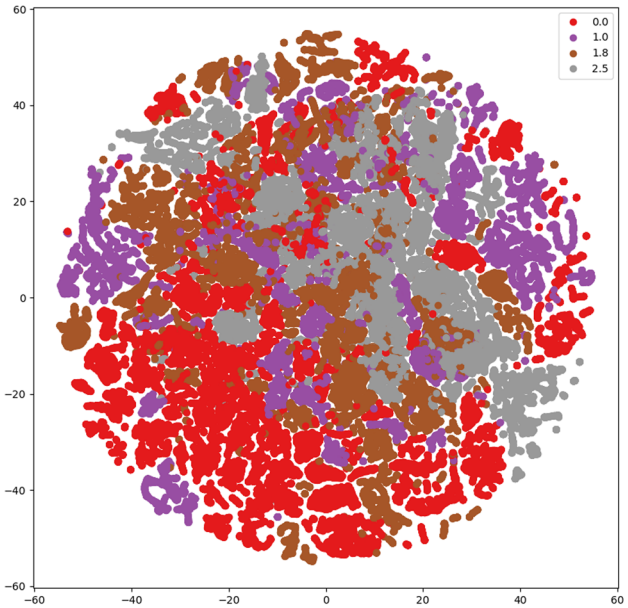

Fig. 7 T-SNE plot for the best performing 8 features mapped to 2D plane

6.1.3 Gaussian mixture regression

Gaussian mixture regression (GMR) [9, 12] is a modified version of Gaussian mixture modeling (GMM) used for regression. GMR is a probabilistic approach that assumes that all the data points in the input $\times$ output space can be effectively represented by a finite number of Gaussian mixtures. As it deals with probabilistic distributions rather than functions, it can model multi-modal mappings. A brief overview of training and prediction details for GMR is given below. Readers are encouraged to go through [12, 47] for further details.

The training for GMR is done by fitting a Gaussian mixture model (GMM) over the feature set F (Eq. 2), in an unsupervised format using the EM algorithm. There is no distinction between input $\mathbf{x}_n$ and target $\mathbf{y}_n$; hence, they can be concatenated into one vector $\mathbf{z}_n = [\mathbf{x}_n^T \mathbf{y}_n]^T$. The GMM represents a weighted sum of E Gaussian functions as a model of the probability density function of the vector $\mathbf{z}_n$, shown in Eq. 3.

$$p(z_n) = \sum_{e=1}^{E} \pi_e \mathcal{N}(\mathbf{z}_n; \mu_e, \sigma_e), \quad \text{with} \sum_{e=1}^{E} \pi_e = 1 \tag{3}$$

For inference, with regression we are interested in predicting $\hat{\mathbf{y}} = E(\mathbf{y}|\mathbf{x})$, i.e., the expected value of $\mathbf{y}$ given $\mathbf{x}$. For this purpose, μ_e and σ_e can be separated into input and output

components as follows:

$$\mu_e = [\mu_{e,X}^T, \mu_{e,Y}^T]; \quad \sigma_e = \begin{bmatrix} \sigma_{e,X} & \sigma_{e,XY} \\ \sigma_{e,YX} & \sigma_{e,Y} \end{bmatrix} \tag{4}$$

Given the decomposition in Eq. 4, the expected value of $\mathbf{y}$ given $\mathbf{x}$ can be calculated by,

$$\hat{\mathbf{y}} = \sum_{e=1}^{E} h_e(\mathbf{x})(\mu_{e,Y} + \sigma_{e,YX}\sigma_{e,X}^{-1}(\mathbf{x} - \mu_{e,X})); \tag{5}$$

where

$$h_e(\mathbf{x}) = \frac{\pi_e \mathcal{N}(\mathbf{x}; \mu_{e,X}, \sigma_{e,X})}{\sum_{l=1}^{E} \pi_l \mathcal{N}(\mathbf{x}; \mu_{l,X}, \sigma_{l,X})} \tag{6}$$

Due to flexibility in the intrinsic nature of probabilistic models, as they are uncertainty-aware and can represent complex problems effectively, we propose to use GMR for modeling the regression problem across various subjects. The experiments, as shown in Sect. 6.3, show that GMR performs well.

6.1.4 Multilayer perceptron regression

A multilayer perceptron (MLP) is a type of artificial neural network (ANN) that is popular due to its ability to learn and recognize complex (non)linear patterns in data. It is a supervised algorithm that is made up of several interconnected layers of neurons, each layer processes and alters the input to conform to an output.

The deterioration (i.e., weight) estimation problem tends to not generalize over all the subjects, i.e., over-fitting to a subset of subjects in training. Thus, a joint loss function is used that includes both lasso ($\mathcal{L}_1$) and ridge ($\mathcal{L}_2$) regularization. If the ground truth label (i.e., weight) is y, and $\hat{y}$ is the regression predicted output, then Eq. 7 shows the loss function. The feature set F (Eq. 2) was used for training.

$$\mathcal{L} = \alpha \|y - \hat{y}\|_2^2 + (1 - \alpha)|y - \hat{y}| \tag{7}$$

where α was set to 0.5.

6.2 Deep features-based regression

Deep features are defined as high-level representations of data learned by deep neural networks (DNN) that capture complex patterns and relationships in data. Deep features possess several advantages over handcrafted features or shallow representations. One key benefit is their automatic inference from the data, allowing the network to dynamically adjust and adapt to the specific task.

To demonstrate generalized regression with deep features, we used a Spatial–Temporal Graph Convolutional Network (ST-GCN) [52]. ST-GCN was chosen for this task as it was the best action recognition algorithm for EatSense, as evidenced in [40].

6.2.1 ST-GCN

When using ST-GCN [52], given the sequence of the body joints (3D in our case), a spatial–temporal graph is constructed with joints as graph nodes, inter-joint connections, and temporal connections (e.g., joint j at time t and $t+1$) as graph edges. By applying spatial–temporal graph convolution operations to the input data, high-level feature maps are generated. Subsequently, a classification head is employed to perform the classification task.

The same approach was used for extracting high-level features. The specific problem here required regression instead of classification. Therefore, two important modifications were made to the ST-GCN framework. Firstly, the classification head was replaced by a regression head. Secondly, the loss function was replaced by the mean-squared error, as described in Eq. 8.

$$\mathcal{L} = \|y - \hat{y}\|_2^2 \tag{8}$$

6.3 Experiments

As mentioned earlier, the experiments for generalized regression are divided into two sub-experiments: handcrafted feature-based regression and deep features-based regression. Each of these sub-experiments has a prior step of hyperparameter tuning. The sub-experiments along with their hyperparameter selection methods are discussed below.

6.3.1 Hyperparameter tuning

The most important hyperparameter for GMR is the number of Gaussians E used to represent the input $\times$ output space effectively. An iterative approach that alternates between searching for E and running 26-vs-1 cross-validation across subjects was used.

In 26-vs-1 cross-validation, 26 subjects were used in the training and validation, and 1 was left out for testing. This was repeated for all 27 subjects, with average results reported. Each set contains different subjects. Searching for the best E used Bayesian optimization to find the configuration that has the minimum mean-squared error across subjects between the ground truth and predicted labels.

The hyperparameters in MLP include the number of layers, neurons in each layer, learning rate, drop-out rate, and batch size. They were chosen empirically. For MLP, similar to GMR, only features selected with the criteria mentioned

in Sect. 6.1.1 were used. Hyperparameters for ST-GCN such as learning rate and others were also chosen empirically.

The experimental question is: How accurately can the amount of weight worn by the subject be estimated (as a proxy for modeling deterioration in elderly eaters)?

6.3.2 Estimating the weight level using regression

After selecting the best configurations, leave-one-out cross-validation was used for measuring the average mean-squared errors (MSE) and actual error for GMR, MLP, LightGBM, and ST-GCN regression. In the leave-one-out approach, the model is trained on all of the available data (here 26 subjects) except for one subject, and then the model's performance is evaluated on the left-out subject. This process is repeated for all subjects, and the overall performance of the model is the average performance across all subjects.

Each of the two sub-experiments used only the 2 most distinctive micro-movements (actions), i.e., 'move hand toward mouth' and 'move hand from mouth.' These 2 actions were chosen because they are the ones that seem most likely to be impacted by varying weights because they involve working against or with gravity. For MLP, LightGBM, and GMR, a frame-by-frame setting of the features was used, whereas for training ST-GCN, vectors containing the 3D poses of one full action each were used to extract feature maps. Afterward, these feature maps go through regression head and predict the weights. The regression models for each subject are used to predict the weight worn by the subject.

To demonstrate the performance of the proposed regression model, we present both visual and quantitative results. Figure 8 shows the predictions of the 27 different models trained using the leave-one-out strategy. Each curve is the output of the one subject who was not involved in the training process. Since the test set comprises multiple instances of the micro-movements, i.e., every subject moves the hand to and away from the mouth multiple times in one eating session, hence these predictions are averaged over time. The solid-colored line represents this mean and the shading around it shows the ± 1 standard deviation of the predictions. For 'summary' purposes, we fit a RANSAC [11] linear regression model across the predicted weights of all 27 of the regression models.[3] In Fig. 8, the black solid line represents the RANSAC linear regression fit line across the predicted weights and the black dashed line illustrates the perfect correlation between the predicted and ground truth weights.

To analyze quantitatively, results are provided using two measures: mean-squared error (MSE) and actual error. The MSE is the ($\mathcal{L}_2$)-norm of the difference between predicted and true values. Likewise, the actual error is the average of the difference between predicted and true values, indicating the deviation in kilograms from the actual weight. Equations 9 and 10 estimate the actual error. Results are given in Table 2.

$$M_p = \frac{1}{N_p} \sum_{n=1}^{N_p} (\text{predicted}_{p,n} - \text{true}_{p,n}) \tag{9}$$

where M_p is the actual error of pth subject in a set of 27 subjects, i.e., $p \in (1, \ldots, 27)$. N_p are the total number of samples in the test set for each person p. So, the overall mean across all subjects is given by,

$$\text{mean} = \frac{1}{27} \sum_{p=1}^{27} M_p \tag{10}$$

6.4 Discussion of results

Both MLP and ST-GCN can handle a wide range of data distributions and excel in different contexts. For example, ST-GCN is specialized for tasks that involve both spatial and temporal dimensions, whereas an MLP can effectively model intricate nonlinearities in high-dimensional data. GMR on the other hand employs a probabilistic approach and models data distributions as combinations of Gaussian mixtures. LGBM works as an ensemble of decision trees and is suitable for tasks where exploitation of high-dimensional feature space is required.

Figure 8 visually compares the effectiveness of three hand-crafted feature-based methods—GMR, MLP and LightGBM, and deep feature-based ST-GCN—using line plots that compare the predicted weights to the ground truth. The more closely the predicted weights (solid black line) align with the actual values (dashed black line), the better the regression model performs.

When examining the results depicted in the top-left figure, it is clear that GMR performs well as there is a noticeable upward trend in the plot, indicating a good prediction of weights (0, 1, 1.8, and 2.4 kg). In contrast, the top-right (MLP) and bottom-left (LightGBM) figures suggest that these models do not generalize as well on the data, as they have a weaker correlation between ground truth and predicted values. The figure on the bottom-right (ST-GCN regressor) clearly shows that the model does not fit properly on the data. This could potentially be due to two reasons, (1) the insufficient temporal context and limited discriminative features as the micro-movements under consideration span over less than 10 frames or (2) insufficient data for training a regression model with only two micro-movements.

When comparing these methods quantitatively, GMR performs better, as evidenced by the average MSE displayed in Table 1. GMR achieved a mean-squared error of 0.53, lower than MLP, LightGBM, and ST-GCN.

[3] RANSAC is a technique that estimates the model parameters by randomly sampling the observed data and hence is robust to outliers.

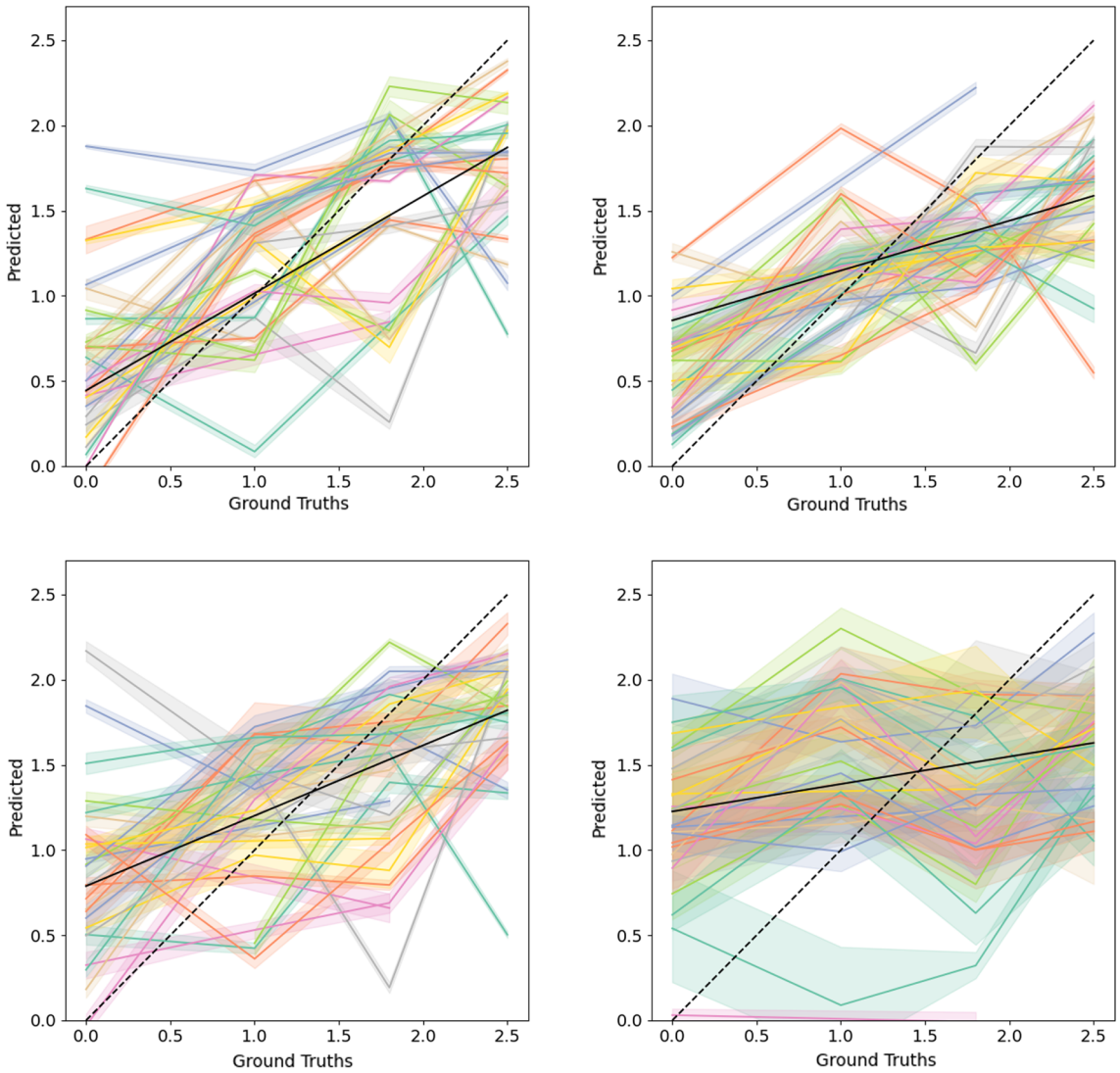

Fig. 8 The plots show the predicted weight versus the ground truth weight. The dashed black line illustrates perfect correlation, and the solid line is the least square fit of the data shown in color. The four regressors evaluated are GMR (top-left), MLP (top-right), LightGBM (bottom-left), and ST-GCN Regression (bottom-right). Each colored curve corresponds to the result of an individual leave-one-out model. Since there are several frames or clips for each micromovement in the test set, the solid-colored curves represent the average of these predictions, while the shading surrounding each curve indicates the range of one standard deviation (color figure online)

In real scenarios, it is unlikely to have data from various stages of deterioration to train a model. Instead, one would have to use one of the generic regression models trained in Sect. 6.3.2. Therefore, relying solely on MSE to quantify the error may seem to be complicated or not intuitive in a physical sense and may not be the most appropriate metric for selecting the best model. To address this, Table 2 presents the actual error (each row estimated by $\frac{1}{N} * \sum_{n=1}^{N}(\text{predicted} - \text{true})$), which indicates the average amount in kilograms that the predictions are off. The table shows that the mean difference for GMR is around 19 gs, with the lowest standard deviation of 0.233. On the other hand, ST-GCN has the lowest mean, with a comparably high standard deviation.

The T-SNE visualization presented in Fig. 7 illustrates that the data have multiple modes, and we anticipate more distinguishable boundaries when considering 8 dimensions. Intuitively, Gaussian mixture regression (GMR) excels in this scenario by representing each mode with its own Gaussian component and clustering data points, rather than attempting to fit a single line or curve across all data. Consequently, GMR demonstrates superior capability in modeling the underlying distributions compared to alternative regression methods.

Table 1 Mean squared error for GMR, MLP, LightGBM and ST-GCN as a result of Leave-one-out regression. The last row shows the average of these errors. Lower values are better, and GMR has the best average performance. Here, bold indicates the best performing approach for each of the 27 models, and the average

S#	GMR	MLP	LightGBM	ST-GCN
0	0.977	0.680	**0.658**	1.540
1	**0.691**	1.856	0.724	1.274
2	**0.189**	1.369	0.845	1.160
3	**0.805**	1.010	1.382	1.210
4	**0.592**	1.363	0.669	0.705
5	**0.404**	1.269	0.859	0.961
6	0.643	0.939	**0.396**	0.663
7	**0.291**	0.581	0.618	0.708
8	0.613	**0.519**	1.674	1.299
9	**0.398**	1.190	0.760	1.069
10	0.931	1.235	1.229	**0.872**
11	**0.597**	0.787	0.738	0.703
12	**0.635**	1.275	0.544	0.975
13	0.629	0.833	**0.420**	1.172
14	0.788	0.627	**0.345**	0.961
15	0.760	0.884	1.279	**0.750**
16	**0.288**	0.432	0.433	1.034
17	**0.598**	0.631	0.629	0.910
18	0.599	0.463	**0.383**	1.290
19	**0.140**	0.313	0.127	0.967
20	**0.327**	0.887	0.329	0.989
21	0.586	1.260	**0.442**	0.814
22	0.284	0.371	**0.177**	0.685
23	**0.328**	0.395	0.452	1.120
24	**0.645**	1.467	0.810	1.327
25	**0.337**	0.538	0.852	1.041
26	0.267	0.834	**0.258**	0.872
Avg	**0.531**	0.889	0.668	1.003

Table 2 Actual error for GMR, MLP, LightGBM and ST-GCN as a result of Leave-one-out regression. The last row shows the mean of these errors. Here, bold indicates the best performing approach for each of the 27 models, and the average. Lower values are better (Values closer to zero are the best.)

S#	G MR	M LP	LightGBM	ST-GCN
0	− 0.256	− 0.186	− **0.164**	− 0.912
1	0.470	0.641	**0.464**	1.019
2	**0.179**	0.898	0.755	0.746
3	− 0.391	0.144	− **0.017**	0.055
4	− 0.259	− 0.296	− **0.064**	− 0.064
5	0.125	0.127	0.072	**0.005**
6	− 0.346	− 0.233	− **0.009**	0.142
7	− 0.153	− 0.328	− 0.093	− **0.009**
8	− 0.422	− 0.337	− **0.240**	0.817
9	− **0.345**	− 0.496	− 0.612	− 0.801
10	0.187	− 0.516	− **0.183**	− 0.222
11	− 0.180	− **0.082**	− 0.542	0.399
12	− 0.246	− 0.497	**0.049**	− 0.181
13	− **0.053**	− 0.356	− 0.256	− 0.124
14	− 0.185	− 0.193	− **0.023**	0.411
15	**0.031**	− 0.234	0.044	− 0.137
16	0.146	− 0.244	− 0.224	− **0.087**
17	0.402	− 0.137	0.526	− **0.001**
18	0.215	− 0.192	0.261	− **0.113**
19	0.141	0.051	**0.027**	0.101
20	− **0.039**	− 0.600	− 0.488	− 0.094
21	**0.009**	− 0.460	0.254	− 0.200
22	0.200	− **0.038**	− 0.049	− 0.141
23	0.134	− 0.183	− 0.225	− **0.120**
24	0.091	− 0.405	0.295	**0.011**
25	− **0.003**	− 0.483	− 0.374	− 0.230
26	**0.031**	− 0.444	0.092	− 0.199
Avg	− 0.019	− 0.188	− 0.026	**0.002**
std	**0.233**	0.333	0.312	0.404

7 Conclusion

In this paper, we presented an analysis of the eating behavior of subjects that includes: modeling the actions involved while eating as a state diagram and methods to quantify performance/decay level. To quantify performance levels while eating, two sets of experiments, i.e., with handcrafted features using uncertainty aware algorithm GMR, with comparisons against MLP and LightGBM, and with deep features-based regression using ST-GCN were conducted.

The results show that GMR performed slightly better compared to other regression models and thus can be used to predict the degree of deterioration (i.e., weight level) of individuals based on a generically trained model (i.e., trained with enough other subject data).

We also presented an extension of the EatSense dataset to four weight levels. Ethical approval was obtained to allow these experiments using healthy human volunteers. In an ideal world, we would also collect long-term data from elderly volunteers to validate the deterioration model; however, this would be highly unethical, as intervention should occur at the first sign of deterioration. Hence, the experiments presented here are limited to using weights with healthy volunteers.

References

1. Alonso, A.C., Luna, N.M., Dionísio, F.N., et al.: Functional balance assessment. Medicalexpress **1**, 298–301 (2014)
2. Amsaprabhaa, M., et al.: Multimodal spatiotemporal skeletal kinematic gait feature fusion for vision-based fall detection. Expert Syst. Appl. **212**(118), 681 (2023)
3. Bartlett, K.A., Camba, J.D.: An RGB-D sensor-based instrument for sitting balance assessment. Multimed. Tools Appl. **82**, 27245–27268 (2023)
4. Barzegar Khanghah, A., Fernie, G., Roshan Fekr, A.: Design and validation of vision-based exercise biofeedback for telerehabilitation. Sensors **23**(3), 1206 (2023)
5. Beard, J., Biggs, S., Bloom, D.E., et al.: Global population ageing: peril or promise? Tech. rep., Program on the Global Demography of Aging (2012)
6. Berg, K.: Balance and its measure in the elderly: a review. Physiother. Can. **41**(5), 240–246 (1989)
7. Blomqvist, M., Luhtanen, P., Laakso, L.: Validation of a notational analysis system in badminton. J. Hum. Mov. Stud. **35**(3), 137–150 (1998)
8. Carnevale, A., Longo, U.G., Schena, E., et al.: Wearable systems for shoulder kinematics assessment: a systematic review. BMC Musculoskelet. Disord. **20**(1), 1–24 (2019)
9. Fabisch, A.: gmr: Gaussian mixture regression. J. Open Source Softw. **6**(62), 3054 (2021). https://doi.org/10.21105/joss.03054
10. Filippeschi, A., Schmitz, N., Miezal, M., et al.: Survey of motion tracking methods based on inertial sensors: a focus on upper limb human motion. Sensors **17**(6), 1257 (2017)
11. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
12. Ghahramani, Z., Jordan, M.: Supervised learning from incomplete data via an EM approach. In: Cowan, J., Tesauro, G., Alspector, J. (eds.) Advances in Neural Information Processing Systems, vol. 6. Morgan-Kaufmann, Burlington (1993)
13. Gill, J., Allum, J., Carpenter, M., et al.: Trunk sway measures of postural stability during clinical balance tests: effects of age. J. Gerontol. A Biol. Sci. Med. Sci. **56**(7), M438–M447 (2001)
14. Grosshauser, F.J., Kiesswetter, E., Torbahn, G., et al.: Reasons for and against nutritional interventions: an exploration in the nursing home setting. Geriatrics **6**(3), 90 (2021)
15. Horak, F.B.: Clinical assessment of balance disorders. Gait Posture **6**(1), 76–84 (1997)
16. Hukkelås, H., Mester, R., Lindseth, F.: Deepprivacy: a generative adversarial network for face anonymization. In: International symposium on visual computing, Springer, pp 565–578 (2019)
17. Iosifidis, A., Marami, E., Tefas, A., et al.: The MOBISERV-AIIA eating and drinking multi-view database for vision-based assisted living. J. Inf. Hiding Multimed. Signal Process. **6**(2), 254–273 (2015)
18. Kanade, A., Sharma, M., Muniyandi, M.: Tele-EvalNet: a low-cost, teleconsultation system for home based rehabilitation of stroke survivors using multiscale CNN-ConvLSTM architecture. In: European Conference on Computer Vision, pp. 738–750. Springer (2023)
19. Ke, G., Meng, Q., Finley, T., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: Guyon, I., Luxburg, U.V., Bengio, S., et al. (eds.) Advances in Neural Information Processing Systems, vol. 30. Curran Associates Inc., New York (2017)
20. Khattar, V., Hathiram, B.: The clinical test for the sensory interaction of balance. Int. J. Otorhinolaryngol. Clin. **4**, 41–45 (2012)
21. Kyritsis, K., Diou, C., Delopoulos, A.: Modeling wrist micromovements to measure in-meal eating behavior from inertial sensor data. IEEE J. Biomed. Health Inform. **23**(6), 2325–2334 (2019)
22. Li, L., Martin, T., Xu, X.: A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. Appl. Ergon. **87**(103), 138 (2020)
23. Li, Z., Huang, Y., Cai, M., et al.: Manipulation-skill assessment from videos with spatial attention network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp. 0–0 (2019)
24. Majumder, S., Aghayi, E., Noferesti, M., et al.: Smart homes for elderly healthcare-recent advances and research challenges. Sensors (2017). https://doi.org/10.3390/s17112496
25. Manghisi, V.M., Uva, A.E., Fiorentino, M., et al.: Real time RULA assessment using Kinect v2 sensor. Appl. Ergon. **65**, 481–491 (2017)
26. Martin, J., Regehr, G., Reznick, R., et al.: Objective structured assessment of technical skill (OSATS) for surgical residents. Br. J. Surg. **84**(2), 273–278 (1997)
27. Meng, L., Chen, M., Li, B., et al.: An inertial-based upper-limb motion assessment model: performance validation across various motion tasks. IEEE Sens. J. **23**(7), 7168–7177 (2023)
28. Merck, C., Maher, C., Mirtchouk, M., et al.: Multimodality sensing for eating recognition. ACM (2016). https://doi.org/10.4108/eai.16-5-2016.2263281
29. Nalci, A., Khodamoradi, A., Balkan, O., et al.: A computer vision based candidate for functional balance test. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp. 3504–3508 (2015)
30. Nayak, G.K., Kim, E.: Development of a fully automated RULA assessment system based on computer vision. Int. J. Ind. Ergon. **86**(103), 218 (2021)
31. Nguyen, T.N., Huynh, H.H., Meunier, J.: 3d reconstruction with time-of-flight depth camera and multiple mirrors. IEEE Access **6**, 38106–38114 (2018). https://doi.org/10.1109/ACCESS.2018.2854262
32. Okamoto, K., Yanai, K.: GrillCam: a real-time eating action recognition system. In: International Conference on Multimedia Modeling. Springer, pp. 331–335 (2016)
33. Onofri, L., Soda, P., Pechenizkiy, M., et al.: A survey on using domain and contextual knowledge for human activity recognition in video streams. Expert Syst. Appl. **63**, 97–111 (2016)
34. Ortells, J., Herrero-Ezquerro, M.T., Mollineda, R.A.: Vision-based gait impairment analysis for aided diagnosis. Med. Biol. Eng. Comput. **56**(9), 1553–1564 (2018)
35. Oshita, M., Inao, T., Ineno, S., et al.: Development and evaluation of a self-training system for tennis shots with motion feature assessment and visualization. Vis. Comput. **35**(11), 1517–1529 (2019)
36. Paiement, A., Tao, L., Hannuna, S., et al.: Online quality assessment of human movement from skeleton data. In: British Machine Vision Conference. BMVA Press, pp. 153–166 (2014)
37. Parmar, P., Morris, B.T.: What and how well you performed? A multitask learning approach to action quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 304–313 (2019)

38. Pauly, L., Stehle, P., Volkert, D.: Nutritional situation of elderly nursing home residents. Z. Gerontol. Geriatr. **40**(1), 3–12 (2007)
39. Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature selection. Pattern Recognit. Lett. **15**(11), 1119–1125 (1994). https://doi.org/10.1016/0167-8655(94)90127-9
40. Raza, M.A., Chen, L., Li, N., et al.: EatSense: human centric, action recognition and localization dataset for understanding eating behaviors and quality of motion assessment. Image Vis. Comput. **137**, 104762 (2023). https://doi.org/10.1016/j.imavis.2023.104762
41. Redfoot, D., Feinberg, L., Houser, A.N.: The Aging of the Baby Boom and the Growing Care Gap: A Look at Future Declines in the Availability of Family Caregivers. AARP Public Policy Institute, Washington, DC (2013)
42. Ren, Y., Lin, C., Zhou, Q., et al.: Effectiveness of virtual reality games in improving physical function, balance and reducing falls in balance-impaired older adults: a systematic review and meta-analysis. Arch. Gerontol. Geriatr. **108**, 104924 (2023)
43. Rolland, Y., Czerwinski, S., Van Kan, G.A., et al.: Sarcopenia: its assessment, etiology, pathogenesis, consequences and future perspectives. J. Nutr. Health Aging **12**, 433–450 (2008)
44. Rosenberg, I.H.: Sarcopenia: origins and clinical relevance. J. Nutr. **127**(5), 990S-991S (1997)
45. Rouast, P.V., Heydarian, H., Adam, M.T., et al.: OREBA: a dataset for objectively recognizing eating behavior and associated intake. IEEE Access **8**, 181955–181963 (2020)
46. Shen, Y., Salley, J., Muth, E., et al.: Assessing the accuracy of a wrist motion tracking method for counting bites across demographic and food variables. IEEE J. Biomed. Health Inform. **21**(3), 599–606 (2016)
47. Stulp, F., Sigaud, O.: Many regression algorithms, one unified model: a review. Neural Netw. **69**, 60–79 (2015)
48. Tufano, M., Lasschuijt, M., Chauhan, A., et al.: Capturing eating behavior from video analysis: a systematic review. Nutrients **14**(22), 4847 (2022)
49. Vuckovic, G., Dezman, B., Pers, J., et al.: Motion analysis of the international and national rank squash players. In: ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005, pp. 334–338. IEEE (2005)
50. Wang, J., Liu, Z., Wu, Y., et al.: Mining actionlet ensemble for action recognition with depth cameras. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1290–1297. https://doi.org/10.1109/CVPR.2012.6247813 (2012)
51. Yadav, R.K., Neogi, S.G., Semwal, V.B.: A computational approach to identify normal and abnormal persons gait using various machine learning and deep learning classifier. In: Machine Learning, Image Processing, Network Security and Data Sciences: 4th International Conference, MIND 2022, Virtual Event, January 19–20, 2023, Proceedings, Part I, pp. 14–26. Springer (2023)
52. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: Proceedings of the AAAI Conference on Artificial Intelligence (2018)
53. Yang, C., Kerr, A., Stankovic, V., et al.: Human upper limb motion analysis for post-stroke impairment assessment using video analytics. IEEE Access **4**, 650–659 (2016)
54. Yang, Z.: An efficient automatic gait anomaly detection method based on semisupervised clustering. Comput. Intell. Neurosci. **2021**, 8840156 (2021)
55. Yordanova, K., Lüdtke, S., Whitehouse, S., et al.: Analysing cooking behaviour in home settings: towards health monitoring. Sensors **19**(3), 646 (2019)
56. Zoidi, O., Tefas, A., Pitas, I.: Exploiting the SVM constraints in NMF with application in eating and drinking activity recognition. In: 2013 IEEE International Conference on Image Processing, pp. 3765–3769. https://doi.org/10.1109/ICIP.2013.6738776(2013)

Muhammad Ahmed Raza received his BS (Electrical Engineering, Pakistan Institute of Applied Sciences (PIEAS), 2016), MS (Electrical Engineering, Air University Pakistan, 2018), and currently, he is doctoral student in the School of Informatics, The University of Edinburgh. He has worked as a research assistant at Air University (2016-2019) and as a research associate in the Swarm Robotics laboratory (University of Engineering and Technology, Pakistan, 2020-2021). His research interests broadly include behavioral assessment (AI for healthcare), object detection, and action recognition frameworks.

Robert B. Fisher FIAPR, FBMVA received a BS (Mathematics, California Institute of Technology, 1974), MS (Computer Science, Stanford, 1978) and a PhD (Edinburgh, 1987). Since then, Bob has been an academic at Edinburgh University, including being College Dean of Research. He has been the Education and Industrial Liaison Committee chairs for the Int. Association for Pattern Recognition, and is currently the association Treasurer. His research covers topics mainly in high-level computer vision and 3D and 3D video analysis, focusing on reconstructing geometric models from existing examples, which contributed to a spin-off company (DI4D). The research has led to 5 authored books and 300+ peer-reviewed scientific articles or book chapters. He has developed several online computer vision resources, with over 1 million hits. Most recently, he has been the coordinator of EC projects 1) acquiring and analyzing video data of 1.4 billion fish from over about 20 camera-years of undersea video of tropical coral reefs and 2) developing a gardening robot (hedge-trimming and rose pruning). He is a Fellow of the Int. Association for Pattern Recognition and the British Machine Vision Association.