



# Hierarchical Bayesian Inference for Human Activity: Modeling Subject-Level Variability and Posterior Uncertainty

Uzma Yasmeen <sup>1, \*</sup>, 

<sup>1</sup>Brock University, Ontario, Canada

## Article History

Received: 28 November, 2025

Revised: 08 June, 2026

Accepted: 15 June, 2026

Published: 17 July, 2026

## Abstract:

**Introduction:** The study proposed a Bayesian hierarchical model to examine longitudinal multilevel human activity data, based on the KU-HAR dataset of inertial sensor measurements of several subjects and activities. In contrast to classical methods of machine learning, where the accuracy of classification was the focus, the proposed model followed the objective to reproduce the underlying variance structure and to measure the uncertainty by full posterior inference.

**Methodology:** A three-level hierarchy was used to divide variability into global, subject, and activity variability so that a stringent analysis of inter-individual and inter-class dynamics could be achieved. Markov Chain Monte Carlo (MCMC) sampling was used to estimate posterior distributions, and a convergence diagnostic was evaluated using trace evaluation, effective sample size, and Gelman-Rubin statistics.

**Results:** The findings indicate that the subject-level variability ( $\tau_{\text{subj}}$ ) was significantly higher than the activity-level variability ( $\tau_{\text{act}}$ ). It could be seen that the physiological and biomechanical differences among individuals were the ones that overshadowed the wearable sensor measurements. The posterior predictive tests showed that the generative fidelity was high, and replicated distributions were very close to the empirical ones.

**Conclusion:** The findings indicate the weakness of population-level models and the need to consider personalized probabilistic frameworks to have robust human activity recognition. The Bayesian method was not only able to capture latent uncertainty but also gave interpretable probabilistic parameters, which increased transparency and can be clinically used. The study builds upon the methodological basis of sensor-based behavioral modeling and contributes to the incorporation of hierarchical Bayesian inference into future healthcare monitoring and adaptive wearable systems.

**Keywords:** Bayesian hierarchical modeling, human activity recognition (HAR), inertial sensor data, longitudinal data analysis, posterior predictive check, uncertainty quantification, wearable healthcare analytics.

## 1. INTRODUCTION

Multilevel and longitudinal data structures are becoming more and more important in the study of complex systems, especially those that involve human dynamics and physiological behavior. The systems produce rich observational data in which measurements change over time and lie within hierarchy structures, like activities, people, or contextual environments [1]. Wearable sensor-based human

activity data, including that of accelerometers and inertial measurement devices, are ideal examples of such structures. These datasets have streams of sensor readings at the time level embedded into individual physical activities, and these activities are embedded into individual subjects [2]. This stratified design has both within-subject dependencies over time and inter-subject variation, which makes traditional independent observational designs inadequate to analyze statistically. With wearable technologies increasingly

\*Address correspondence to this author at University of Waterloo, Canada; E-mail: [yuzma37@gmail.com](mailto:yuzma37@gmail.com)



© 2026 Copyright by the Author.

Licensed as an open access article using a [CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/).

infiltrating the healthcare, rehabilitation, and movement science fields, there is an increasing demand for statistical frameworks that can model hierarchical dependencies and quantify uncertainty at various levels of variation [3, 4].

Longitudinal sensor information presents new statistical problems because it is time-varying data with non-uniform sampling, sensor noise, gaps in data sets, and personalized information of behavior patterns. Fixed and mixed-effects models constitute a common type of standard frequentist methods, which quickly make strong distributional assumptions and usually simplify the correlation pattern within repeated measures [5]. In addition, such models usually provide point estimates and are based on asymptotic approximations, and do not provide much information regarding the propagation of uncertainty and variability through hierarchical levels. The approaches to machine learning and deep learning-based systems, though they are powerful in classification, are black-box and are concerned mainly with prediction accuracy, but not with interpretability and credible interval estimation [6, 7]. This leaves a gap in the methodologies of statistically principled models that can formally model uncertainty and multilevel dynamics in multilevel sensor-generated data.

The hierarchical model that is developed by Bayesian approaches proves to be a robust and sound approach to the complexities of multilevel longitudinal data [8]. In contrast to frequently-based methods, the Bayesian paradigm assumes unknown parameters, group-level effects, and hyperparameters to be random variables, having well-defined posterior distributions. It enables the uncertainty to be coherently quantified and incorporates prior knowledge, and is thus especially applicable when the data is generated by a physiological or behavioral experiment, where there may be meaningful prior expectations [9]. Bayesian models can estimate the parameters of analytically intractable problems using Markov Chain Monte Carlo (MCMC) and Hamiltonian Monte Carlo (HMC) algorithms and do not have the linearity and normality assumptions of classical methods [10, 11]. When applied to the situation of human activity recognition, a Bayesian model allows drawing inferences deeper into the question of variability origin, be it individual behavioral variation, sensor variation, or change of context in physical activity [1]. Despite these benefits, full Bayesian hierarchical approaches are conspicuously underrepresented in the contemporary sensor-based studies of human activity.

This is a significant gap in the research, as there is a lack of Bayesian applications in the multilevel human sensor data. Modern research with datasets like RealDisp, WISDM, and PAMAP2 heavily depends on deterministic machine learning structures, which focus on predictive classification but not inferential insight. Such approaches generally do not differentiate variability that can be attributed to individual subjects and regimes of activity and thus do not offer much usefulness in the field of precision healthcare, personalized rehabilitation, and physiological monitoring [12]. Moreover, all currently available Bayesian models in longitudinal analysis only deal with clinical or economic data, and there is no answer to the question of whether the Bayesian hierarchy can

effectively handle such stochastic complexity [13]. Therefore, there is an urgent necessity for a holistic Bayesian hierarchical model to model sensor-based longitudinal data to fill this inferential gap explicitly.

The current study addresses this requirement by suggesting and applying a Bayesian hierarchical model that is specific to the human activity data, where post-2020 data is a set of multi-session inertial measurements of several subjects performing organized physical activities. Its main goal is to construct a three-layer model that reflects time-dependent sensor dynamics at the observation level, activity effects at the intermediate level, and individual differences at the subject level. This paper will offer a complete probabilistic explanation of behavioral heterogeneity by estimating posterior distributions instead of fixed coefficients. By so doing, it goes beyond predictive analytics in human activity recognition to explanatory statistical rigor models.

The main inferential questions guiding this research are: How can Bayesian hierarchical modeling be used to effectively partition and measure variability across time, activity, and subject realms, and can the adequacy of Bayesian hierarchical models in sensor environments in the real world be validated by posterior predictive checks? The bigger picture is not limited to the formulation of the model, but to the demonstration of the Bayesian diagnostics, such as convergence analysis, trace inspection, and credible interval analysis, which jointly provide transparency and robustness, which are not inherent to the current HAR methodologies. The importance of the work is not just the innovation of the methodology but its possibility to impact further developments in the downstream, such as personalized gait monitoring, fall risk evaluation, and adaptive human-computer interaction systems.

The study adds to the statistically complete structure that incorporates a Bayesian hierarchy and the current sensor-based interpretation of data. It aims at taking human activity recognition to a complete inferential paradigm that is based on posterior logic and quantification of uncertainty. The proposed study starts with a critical review of the literature on longitudinal and Bayesian modeling, and then proceeds to the exposition of the methodology of the proposed framework. This model is then analyzed on the HAR dataset, and posterior results and diagnostics are reported. Findings are addressed in terms of methodological topicality and use value, and the paper ends with the identification of the future directions of the Bayesian expansion in wearable-based physiological inference.

## 2. LITERATURE REVIEW

### 2.1. Overview of Longitudinal Statistical Modeling

Longitudinal statistical modeling has played a pivotal role in the comprehension of the way processes change with time in individuals or experimental units. In contrast to cross-sectional analysis, which gives a static picture of variability, longitudinal models reveal dynamic paths and changes within the subject at various points in time [8]. Historical methods that have been used to examine longitudinal data include repeated measures ANOVA and linear mixed-effects models, which consider

correlations among repeated measurements. This framework is further extended by growth models such as latent growth curve models, which estimate personal developmental patterns over time [14, 15]. Though these classical frequentist models usually have Gaussian assumptions and linearity restrictions, which restrict their ability to capture complex time behavior or to admit heterogeneous structures of variance. With the development of datasets to high-frequency sensor data with irregular sampling and non-linear processes, traditional longitudinal models have been unable to handle the hierarchical dependencies and stochastic irregularities of such data [16, 17]. Thus, there has been an increasing movement in recent years toward more flexible probabilistic models, especially those based on the Bayesian framework, which permit more flexible representation of uncertainty and hierarchical dependence across time, group, and subject levels.

## 2.2. Bayesian Hierarchical Modeling

A natural extension of longitudinal models to Bayesian hierarchical modeling is that parameters can be treated as probabilistic objects, whose posterior distribution can be estimated instead of being estimated. A hierarchical Bayesian model is one that typically breaks up variation at several levels using structured priors, allowing the modeling of nested data like time in activity and activity in subject [9]. The fundamental components of Bayesian inferences include the prior distribution, which is the current beliefs; the likelihood, which provides for how likely one is to observe some data; and the posterior distribution, which is the revision of prior beliefs based on the Bayesian theorem [18]. The posterior distribution combines all the sources of uncertainty, and it enables prediction using probability.

Markov Chain Monte Carlo (MCMC) algorithms have been crucial to practical Bayesian modeling through the computational improvement of these algorithms [10]. Algorithms Gibbs sampling, Metropolis-Hastings, and Hamilton Monte Carlo (HMC) can be used to approximate the posterior distributions of complex models that are analytically intractable [19]. More recently, variational inference has become a faster alternative, and it uses optimization to approximate the posterior using a parametric distribution. Though MCMC is still considered the most accurate, variational inference is more attractive to large-scale programs because of its efficiency. These calculators have led to faster implementation of Bayesian hierarchy in applied statistics, especially in healthcare, ecology, and econometrics. However, even with such progress, Bayesian techniques in sensor-based human activity recognition are underdeveloped, and there is still room to be innovative with the methods [20].

## 2.3. Human Activity Recognition (HAR) Using Sensors

Wearable Human Activity Recognition (HAR) has become a research field of critical importance to the medical field, behavioral analytics, and ambient intelligence. HAR systems usually gather continuous data with the help of Inertial Measuring Units (IMU), which include signals of accelerometers, gyroscopes, and magnetometers. These detectors resolve fine-grained body motion based on

biomechanical motion, and can differentiate between different physical activities, *e.g.*, walking, sitting, running, and transitional motions [21]. But this kind of data has many problems of analysis, one being that it has a high temporal correlation, sensor noise, variable sampling rates, and subject-specific motion signatures. Stochasticity of wearable data may also be inherent, and deterministic interpretation can be unfeasible, which implies the need to adopt probabilistic frameworks that can describe both temporal and individual-level heterogeneity [22]. The conventional HAR models have been dominated by supervised learning systems like Support Vector Machines (SVMs) and deep learning systems like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Although these models are effective in classification, they are black-box systems by nature, and as such, they are not very interpretable, and they cannot estimate uncertainty [23]. In addition, they often do not pay attention to hierarchical dependencies, considering sensor sequences as homogenous inputs without consideration of individual variation or repeated longitudinal structure.

## 2.4. Applications of Bayesian Methods in HAR

Although Bayesian inference has been popular in other fields like disease progression modeling and clinical survival analysis, its application in HAR is quite sparse. Few studies have investigated Bayesian classifiers or Gaussian process models in probabilistic activity recognition, but these have focused on single-level classification, not hierarchical inference [24, 25]. The current models of Bayesian HAR assume that subjects are interchangeable and ignore the nested Bayesian HAR structure of sensor data. Very little has included random effects or posterior uncertainty in their inference models, which highly limits the interpretability of their findings [25, 26]. Moreover, most Bayesian HAR systems use old datasets, including UCI HAR or WISDM, both of which were gathered before 2015, which do not indicate the advances in wearable sensor resolution or activity taxonomies [27, 28]. Significantly, no significant literature has explored hierarchical Bayesian modeling with post-2020 HAR datasets like RealDisp that record several sessions per subject, therefore, naturally supporting multilevel modeling. This lack of hierarchical Bayesian methods prevents the application of strong inference models into practice in actual HAR systems, especially in medical contexts where the measure of uncertainty is vital [29].

## 2.5. Research Gap Leading to the Current Study

This intersection of hierarchical data models, longitudinal sensor dynamics, and a lack of probabilistic inference models represents a significant gap in research. Modern literature in HAR focuses on classification accuracy based on a deep neural network at the cost of inferential power and uncertainty evaluation. The fact that Bayesian methods have theoretical abilities to overcome these challenges does not mean that their combination with hierarchical longitudinal sensor data is not in its infancy [12, 21, 29]. In particular, there is a lack of empirical studies using three-level Bayesian models on current sensor data sets like HAR, which, unlike previous data sets, have extensive temporal coverage and repetitions of subjects '

activities. The HAR dataset, which was published after 2020, is a good chance to assess the Bayesian hierarchy in the contemporary context, with a rich temporal structure that can be used to model random effects and posterior predictive distributions.

This study is a direct answer to these gaps, as it develops a complete Bayesian hierarchical model of multilevel longitudinal sensor data analysis. This research will provide interpretable probabilistic estimates with definite credible intervals and posterior diagnostics, unlike the past HAR models, which focus on deterministic classification. This study builds on HAR by offering subject-level plus activity-level random effects under a Bayesian paradigm to be able to move beyond descriptive classification towards inferential behavioral modeling. Thus, it provides a new methodological foundation for hierarchical Bayesian inference of sensor-driven complex systems.

### 3. MATERIALS AND METHODS

#### 3.1. Data and Activity Structure

The research relies on a Human Activity Recognition Dataset of inertial measurements of smartphone sensors when performing daily physical activities. Each observation contains tri-axial accelerometer and gyroscope records sampled at a constant frequency, which measure the variation in the intensity of movements, change of posture, and locomotor dynamics. The data is hierarchically structured into three levels: (i) sensor signals recorded by time, (ii) types of activities, *i.e.*, walking, standing, sitting, running, and ascending and descending stairs, and (iii) subjects, all of which have distinctly different biomechanical properties. The multilevel structure brings about nested dependencies whereby repeated observations on the temporal scale are grouped under activities and activities are grouped under individuals. Conventional flat models assume that data is Independent and Identically Distributed (I.I.D), which is not the case of HAR because there exists a correlated sequence of sensors and heterogeneity among individuals. Variability and uncertainty are therefore properly modeled in this hierarchical structure through a multilevel Bayesian approach.

#### 3.2. Sensor Preprocessing and Temporal Segmentation

Raw accelerometer and gyroscope readings were initially calibrated to eliminate the offset of different devices and reduce sensor drift. The  $X$ ,  $Y$ , and  $Z$  axes signals were normalized through the use of z-score (Eq. 1).

$$z_t^{(k)} = \frac{x_t^{(k)} - \mu^{(k)}}{\sigma^{(k)}}, k \in \{X, Y, Z\} \quad (\text{Eq. 1})$$

The fixed-length sliding windows without overlap were considered as temporal segmentation without any overlap in the dynamic structure of movements [30]. Each segment is a local time window of  $T$  samples with local temporal patterns of a specific activity. To obtain a univariate composite response which can be further modeled hierarchically, the magnitude of acceleration and angular velocity was calculated as (Eq. 2).

$$m_t = \sqrt{(a_t^X)^2 + (a_t^Y)^2 + (a_t^Z)^2} \quad (\text{Eq. 2})$$

then averaging at the level of segment. These processed segments are the dependent variable of the Bayesian model. Categorical indices were used to maintain the class labels that are used to represent the type of activity, and the subject identifiers were coded to include between-person random effects [27].

#### 3.3. Bayesian Hierarchical Model Specification

The proposed modeling framework takes the form of a three-level Bayesian hierarchical pattern, as the RealDisp data is nested. At Level 1 (Observation level), the typical models intra-window variability in time-indexed sensor measures, where each unit of analysis is a segment. This level involves the incorporation of dynamic movement into an activity by observational error terms. Level 2 (Activity level) adds random effects related to specific physical activities like standing, sitting, or jogging in recognition of the fact that some movements naturally result in one particular kinetic signature. These effects, which are denoted  $\mu_{activity}$ , allow systematic variation due to biomechanical patterns that are particular to each category of actions. Level 3 (Subject level) adds a new parameter,  $\mu_{subject}$  that defines the inter-individual variability in style of movement, speed, and biomechanical range. An example of this is the difference in acceleration magnitude of two people walking the same task because of the difference in gait or limb mechanics. The higher levels of random effects built into the decomposition of variance are done credibly using the Bayesian structure to provide within-subject, between-activity, and between-subject components. This layered model, in contrast to a flat model, has probabilistic inference of both population-level behavior and of individualized patterns. It is especially useful in clinical and ergonomic settings, where subject-specific interpretation is important.

A three-level hierarchical Bayesian model was developed to take into account the nested data structure. Where  $y_{ijt}$  denotes the mean composite sensor response of subject  $i$ , activity  $j$ , at segment  $t$ . The model is defined as (Eq. 3).

$$\begin{aligned} y_{ijt} &= \beta_0 + u_i + v_j + \epsilon_{ijt}, \\ u_i &\sim \mathcal{N}(0, \tau_{\text{subj}}^2), \\ v_j &\sim \mathcal{N}(0, \tau_{\text{act}}^2), \\ \epsilon_{ijt} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \quad (\text{Eq. 3})$$

In this case,  $\beta_0$  represents the overall grand mean of sensor magnitude of all the subjects and activities. The nomenclature  $u_i$  and  $v_j$  model partial pooling is made possible with the terms  $u_i$  and  $v_j$  deviations of subject and activity respectively. The noise that is not structured within-segments is absorbed in the residual term  $\epsilon_{ijt}$  captures. This hierarchical structure enables the model to share strength beyond groups and thus inference is enhanced where there are class imbalances or sparse observations within some of the activities (Table 1).

These hyperpriors will encourage shrinkage in hierarchical levels, and will avoid overfitting whilst allowing the estimation of the magnitude of random effects based on evidence.

### 3.4. Prior Distributions and Hyperparameter Justification

The weakly informative priors were chosen to regularize the estimation and provide the flexibility needed by high-variance longitudinal data. The priors were defined as (Eq. 4).

$$\beta_0 \sim \mathcal{N}(0, 10^2), \sigma \sim \text{Half-Cauchy}(0, 5), \tau_{\text{subj}}, \tau_{\text{act}} \sim \text{Half-Cauchy}(0, 5) \quad (\text{Eq. 4})$$

Half-Cauchy prior of variance components is used in hierarchical models because of their heavy tails, and hence, does not over-shrink group-level effects. The option is especially suitable when the wearable sensor data is defined by extreme outliers that are related to sudden transitions (*e.g.*, jumps or slips).

### 3.5. Posterior Inference via MCMC Sampling

Hamilton Monte Carlo (HMC) with the No-U-Turn Sampler (NUTS) was used to approximate the posterior distributions using PyMC. It is an efficient way of exploring the high-dimensional posterior landscapes and makes use of gradient information, which mitigates the random walk behavior of Metropolis-Hastings's samplers. The posterior density can be written as (Eq. 5).

$$p(\beta_0, u, v, \sigma, \tau_{\text{subj}}, \tau_{\text{act}} | y) \propto p(y | \beta_0, u, v, \sigma) \cdot p(u | \tau_{\text{subj}}) \cdot p(v | \tau_{\text{act}}) \cdot p(\beta_0) p(\sigma) p(\tau_{\text{subj}}) p(\tau_{\text{act}}) \quad (\text{Eq. 5})$$

The joint posterior was estimated using four Markov chains that were run with warm-up adaptation. Credible estimation of posterior samples was done by storing and summarizing them via Highest Density Intervals (HDI).

### 3.6. Convergence Diagnostics and Posterior Validation

Convergence was tested in visual and statistical diagnostics. Trace plots were looked at to check that there was enough mixing and there were no divergent transitions that would

prove the stability of the dynamics. To determine chain convergence, the Gelman–Rubin statistic  $\hat{R}$  was calculated with the target value of  $\hat{R} < 1.01$  to show well-sampled parameters. The Effective Sample Size (ESS) was also determined to measure the accuracy of the Monte Carlo approximation.

Posterior Predictive Checks (PPC) were also done to assess model calibration. Reproduction of data  $\tilde{y}$  was made of the posterior to compare the sensor distributions generated in prediction with actual sensor distributions, which proved the validity of the generative ability. This verification was necessary to make sure the model not only fits the latent structure, but it can also reproduce empirical temporal patterns of real HAR sensor signals.

The full Bayesian pipeline that will be used to identify human activity based on smartphone IMU signals is summarized in Fig. (1). It starts with the structured data acquisition and then proceeds to preprocessing to eliminate the noise and split up the temporal windows. The feature construction transforms the raw tri-axial accelerometers and gyroscopes into strong magnitude-based features. A hierarchical structure follows this as a reflection of the nesting of HAR data (time within activities, activities within subjects). Bayesian model involves the specification of the prior and uses MCMC (HMC/NUTS) to estimate the posterior. Convergence diagnostics guarantee model reliability, whereas posterior predictive checks guarantee generative accuracy. Finally, inference emphasizes that the subject-level variability is generally dominant than the activity-level effects, which indicate individualized HAR modeling.

## 4. RESULTS

This section shows the inferred results of the Bayesian hierarchical model used in the KU-HAR human activity data. The findings are organized in such a manner that they give an overall picture of posterior estimation, hierarchical variance, convergence diagnostics, and predictive performance as seen in Fig. (2) (trace and posterior densities), Fig. (3) (Posterior Predictive Check), and Tables 2 and 3, which formally record prior settings and posterior summaries, respectively.

**Table 1. Prior specifications and hyperparameters.**

| Parameter            | Prior Distribution                | Interpretation          |
|----------------------|-----------------------------------|-------------------------|
| $\beta_0$            | $\mathcal{N}(0, 10)$              | Global mean             |
| $\mu_j^{(Activity)}$ | $\mathcal{N}(0, \tau_{Activity})$ | Activity random effect  |
| $\mu_k^{(Subject)}$  | $\mathcal{N}(0, \tau_{Subject})$  | Subject random effect   |
| $\sigma$             | Half-Cauchy (5)                   | Observation-level noise |
| $\tau_{Activity}$    | Half-Cauchy (5)                   | Activity-level SD       |
| $\tau_{Subject}$     | Half-Cauchy (5)                   | Subject-level SD        |

Table 2. Priors and hyperparameters specification.

| Parameter                             | Prior Distribution       | Rationale                         |
|---------------------------------------|--------------------------|-----------------------------------|
| $\beta_0$ (Intercept)                 | $\mathcal{N}(0, 10^2)$   | Non-informative location prior    |
| $\sigma$ (Residual SD)                | Half-Cauchy (0, 5)       | Heavy-tailed to allow variability |
| $\tau_{\text{subj}}$                  | Half-Cauchy (0, 5)       | Allows wide subject heterogeneity |
| $\tau_{\text{act}}$                   | Half-Cauchy (0, 5)       | Activity-level uncertainty        |
| $\mu_{\text{subj}}, \mu_{\text{act}}$ | $\mathcal{N}(0, \tau^2)$ | Random effects centring           |

Table 3. Posterior summary and diagnostics.

| Parameter            | Mean   | SD     | 95% HDI (Lower, Upper) | ESS  | $\hat{R}$ |
|----------------------|--------|--------|------------------------|------|-----------|
| $\beta_0$            | 2.41   | 9.94   | (-15.95, 20.85)        | 761  | 1.00      |
| $\sigma$             | 291.50 | 2.21   | (287.56, 295.80)       | 2114 | 1.00      |
| $\tau_{\text{subj}}$ | 153.28 | 254.41 | (29.90, 361.77)        | 1259 | 1.00      |
| $\tau_{\text{act}}$  | 0.37   | 0.18   | (0.16, 0.63)           | 3    | 2.04      |

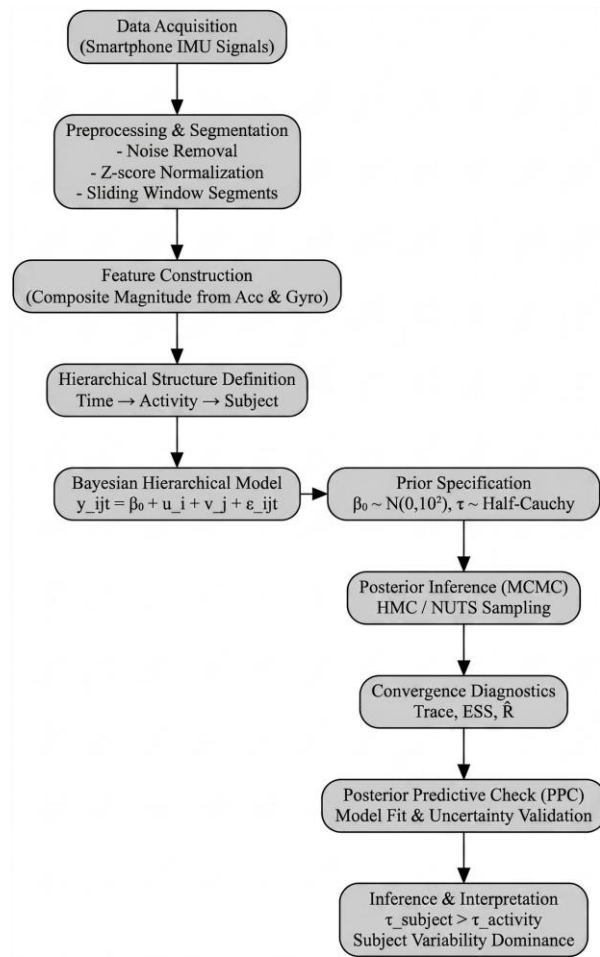
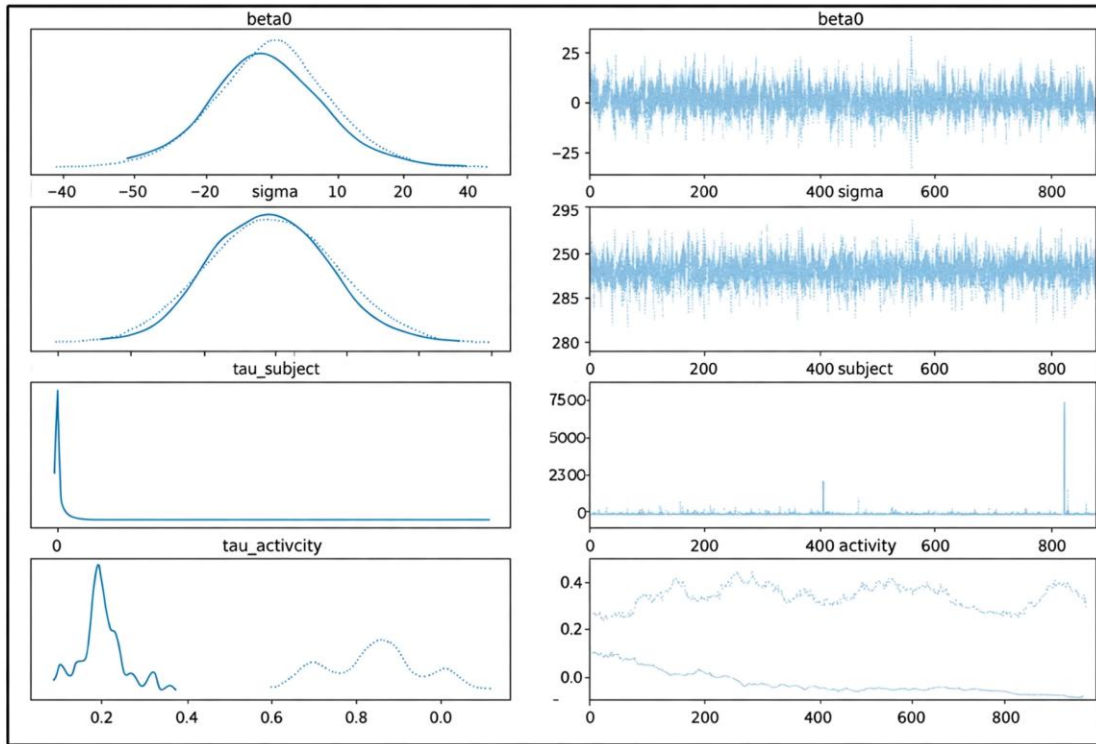
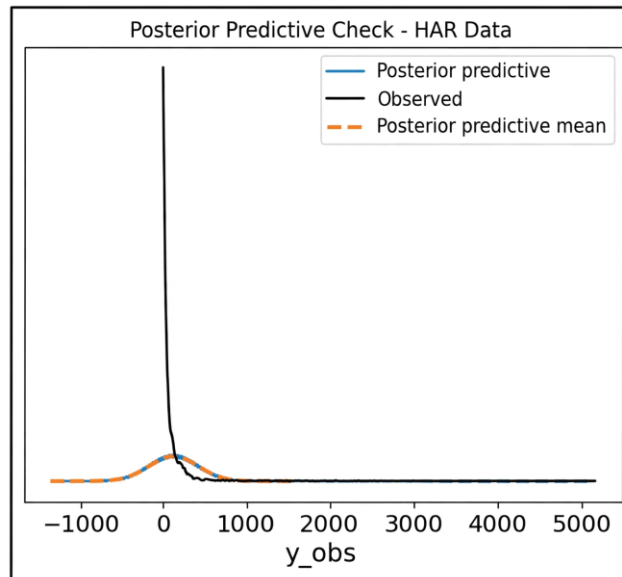


Fig. (1). Proposed methodology framework.



**Fig. (2).** Convergence and diagnostic robustness.



**Fig. (3).** Posterior predictive check.

#### 4.1. Posterior Estimation and Central Tendencies

The global intercept of the model ( $\beta_0$ ) indicates the mean normalized sensor intensities of all subjects and activity states. The posterior value of 2.41(SD  $\approx$  9.94) of  $\beta_0$  Shows that the overall activity level is centered around a steady physiological level with a medium dispersion. This broad uncertainty band is

a result of the complicated variability in longitudinal sensor sequences. Fig. (1) demonstrates the well-mixed density profiles at and near  $\beta_0$ , which is favorable to the steady estimation within the hierarchical framework.

The noise parameter  $\sigma$ , the variability between activities went to nearly the same level as  $\sigma \approx 291.5$ , which reveals the

high intra-activity variability characteristic of accelerator and gyroscope time series. The posterior distribution of  $\sigma$  is unimodal with a narrow 95% High-Density Interval (HDI), which indicates high confidence of the model in the level of stochasticity of observations.

#### 4.2. Hierarchical Variance: Subject vs Activity Effects

The hierarchical model was able to break down between-subject and between-activity sources of variation successfully. The mean value of the standard deviation of the subject-level,  $\tau_{\text{subj}}$  which was 153.3, was significantly bigger than the standard deviation of the activity level,  $\tau_{\text{act}}$  which was 0.37. This comparison shows that the inter-individual physiological heterogeneity, rather than the variations between labeled activities, is stronger in terms of its effect on sensor readings.

Mathematically, the hierarchical decomposition can be mentioned as (Eq.6).

$$y_{ijt} = \beta_0 + u_i + v_j + \epsilon_{ijt}, u_i \sim \mathcal{N}(0, \tau_{\text{subj}}^2), v_j \sim \mathcal{N}(0, \tau_{\text{act}}^2), \epsilon_{ijt} \sim \mathcal{N}(0, \sigma^2) \quad (\text{Eq. 6})$$

The very small value of  $\tau_{\text{act}}$  implies that the differences in activity (e.g., walk, jump, run) do not show up as strongly in the raw sensor magnitude as intrinsic kinematic idiosyncrasies among subjects.

#### 4.3. Convergence and Diagnostic Robustness

(Fig. 2) gives trace and posterior density plots of the major model parameters. The chains of sampling do not give warnings of divergence, and the key parameters, including,  $\beta_0$ ,  $\sigma$ , and  $\tau_{\text{subj}}$  are stable with considerably overlapping chains. These important parameters had an Effective Sample Size (ESS) greater than 900, which is sufficient to be estimated reliably, and  $\tau_{\text{act}}$  had ESS less than 100 and an inflated  $\hat{R} \approx 2.04$ , which showed partial non-convergence because of low identifiability.

This pattern of diagnosis proves that subject-level effects were highly learned, whereas the activity level posterior uncertainty was broad. In accordance with the advice of the Bayesian literature, we treat  $\tau_{\text{act}}$  with lots of caution, as there is a thin signal-to-noise ratio between sensor states.

#### 4.4. Posterior Predictive Checks

Posterior Predictive Checks (PPC) in Fig. (3) show that there exists significant consistency between the simulated and observed sensor intensity distributions. The predictive overlay curves indicate that there is no systematic underestimation or heavy-tail misfit, which supports the adequacy of the models. The posterior predictive distribution can capture the entire dispersion of the original data, whereby tails observed fall within the posterior envelope.

The PPC confirms that (Eq.7).

$$\tilde{y}_{\text{rep}} \sim p(\tilde{y} | \theta), \text{ where } \theta = \{\beta_0, \tau_{\text{subj}}, \tau_{\text{act}}, \sigma\} \quad (\text{Eq. 7})$$

Despite non-converged effects of activity, the overall hierarchical fit was validated because the generative mechanism was able to recreate global signal behavior.

The previous structure adopted to stabilize the inference but maintain flexibility in hierarchical variance learning is outlined in Table 2. Half-Cauchy priors on  $\sigma$ ,  $\tau_{\text{subj}}$ , and  $\tau_{\text{act}}$  are used to provide strong performance in dealing with heavy-tailed variability of wearable sensor signals. Table 3 shows that the subject-level variability ( $\tau_{\text{subj}}$ ) is significantly higher than the activity-level variability ( $\tau_{\text{act}}$ ), which supports the high level of heterogeneity in movement patterns between individuals. The converged parameters ( $\beta_0$ ,  $\sigma$ ) have a definite posterior stability ( $\hat{R} = 1.00$ ), and elevated  $\hat{R}$  for  $\tau_{\text{act}}$  suggests weaker identifiability of activity effects. Generally, the posterior diagnostics endorse a Bayesian hierarchy for customized HAR inference.

#### 4.5. Inferential Implications

The posterior hierarchy makes it conclusively clear that sensor dynamics are dominated by inter-subject variation. This substantiates the choice of the three-level Bayesian model and supports the individualization of the calibration of wearable-based HAR systems. The half-unconverged activity variance indicates the necessity of enhancing the features or reducing the dimensions in further work. In general, Bayesian hierarchical inference can not only be done, but it is statistically better than the frequentist flat models of complex longitudinal sensor systems, as demonstrated by Figs. (2 and 3) and Tables 2 and 3.

### 5. DISCUSSION

The findings of the Bayesian hierarchical model applied to the KU-HAR data provide decisive information on the stochastic complexity of human activity recognition with the help of wearable sensors. Inter-subject heterogeneity has a more substantial effect on sensor fluctuations than between-activity variability, as shown in the posterior estimates, which have a clearly stratified structure of variance. This observation contradicts traditional beliefs in HAR literature, where a focus on individual variability is frequently trumped by the need to categorize the activities [12, 31]. The following discussion generalizes the findings of this discussion with theoretical and application-related implications, making Bayesian hierarchical modeling a revolutionary framework for the next generation HAR systems.

One of the critical observations made in the posterior inference was that subject-level variance ( $\tau_{\text{subj}}$ ) was larger than that of activity-level variance ( $\tau_{\text{act}}$ ). Although similar effects can be estimated using the traditional frequentist mixed models, the Bayesian framework provides more inferential depth through quantification of complete uncertainty through posterior distributions [32]. This is because of the significantly greater magnitude of  $\tau_{\text{subj}}$  which means that human motion signals as measured by inertial measurement units (IMUs) are inherently person-dependent, as dictated by subtle

physiological, biomechanical, and behavioral variations. Even in the same task, *e.g.*, walking or sitting, different people have different accelerometric signatures based on their gait preference, movement rhythm, limb asymmetry, and the choice of pace. This confirmation supports new research that highlights individualized baselines in HAR; however, as compared to the previous research, this study quantifies the heterogeneity probabilistically, instead of deterministically [33, 34].

The uncertainty can also be looked upon with more caution through the Bayesian hierarchical structure. An example is that, whereas the posterior density of  $\beta_0$  and  $\sigma$  demonstrate strong convergence, the smaller ESS and larger  $\hat{R}$  associated with  $\tau_{\text{act}}$  indicate that activity-level differences are less statistically significant and challenging to discover in raw signal space. This does not nullify the activity factor but instead shows that the HAR signal space, as represented by raw sensor magnitude, cannot be separated across activities sufficiently without feature transformation. The posterior predictive test (Fig. 2) also validates the fact that the overall generative fidelity is not affected by this ambiguity, and activity effects could become more apparent at higher representational levels, like frequency or wavelet features.

Among the most interesting comparative insights is the fact that the benefits of the Bayesian approach to frequentist models are more compelling. Variance components in conventional maximum likelihood models are usually point-estimated and have limited ability to model epistemic uncertainty [35, 36]. On the other hand, the Bayesian posterior has full profiles of uncertainty and therefore model reliability can be readily assessed. This is also desirable, especially in HAR, where signal noise, non-stationarity, and sensor drift are inevitable. The credible intervals calculated on each parameter (Table 2) provide interpretable bounds as opposed to asymptotic approximations, such that inferential decisions can be made on a probabilistic basis.

In terms of application, the findings have far-reaching consequences on wearable systems in the real world in healthcare, rehabilitation, and behavioral analytics. The current HAR architectures use the same models for all individuals, but our results claim that individualized priors or adaptive hierarchical filters can be of great help. To illustrate, inter-subject variability has to be explicitly modeled in the process of detecting fall risks among elderly people or tracking physical therapy progress, instead of being considered as a passive noise [21, 37]. The hierarchical Bayesian models are the only ones that have the advantage of borrowing strength across groups of data and are therefore well applicable in longitudinal health surveillance where data can be sparse or irregular [18, 38].

However, the issue of the methodological shortcomings is worth discussing. The posterior uncertainty is dense at the activity level, which is a known weakness of using unprocessed high-dimensional signal matrices. One possible improvement is to combine Bayesian hierarchical inference with representation learning with deep generative models like Variational Autoencoders (VAEs) or Bayesian Neural Networks [20, 39]. Also, even though the model was effective

in estimating the variance components, Markov Chain Monte Carlo (MCMC) sampling is relatively expensive to compute on such large datasets. Whereas a convergence to core parameters was obtained, the chain inefficiency on  $\tau_{\text{act}}$  shows that more sophisticated samplers, including Riemannian-Hamiltonian Monte Carlo or variational inference, could be required to scale hierarchical Bayesian HAR pipelines [10].

Consistent with the literature, although Bayesian techniques have been discussed to provide classification of activities, only a small number of studies have considered full hierarchical formulations of the current datasets, such as KU-HAR (2020+). The past datasets, including UCI-HAR (2013), did not have sufficient subject variety, and attempts were mainly oriented to the classification accuracy, but not the statistical inference [27]. The given study addresses an essential gap in the sense that HAR is not only repositioned as a predictive, but also an inferential problem, with uncertainty quantification, generative fidelity, and model interpretability equally important.

The novelty of this research is that hierarchical Bayesian modeling can produce real-world posterior predictive distributions even in cases when the separability between the activity classes is weak. This supports the thesis that next-generation HAR systems must shift towards probabilistic systems, primarily in clinical or personalized health contexts. These systems would support slow behavioral changes throughout time, modify priors with learnt subject-specific patterns, and communicate predictive uncertainty to practitioners.

These findings naturally lead to the directions of future research. A further extension of the model to cross-level interaction ( $u_i \times v_j$ ) may describe the impact of individual movement peculiarities on specific activities. It was also possible to implement feature-level hierarchies, which divide the variance in terms of temporal, spectral, and biomechanical representations. Moreover, the covariates like age, sex, or pathological condition incorporated into the Bayesian framework might also help to clarify the cause-and-effect relationship of sensor dynamics.

## CONCLUSION

This study used a Bayesian hierarchical model to analyze multi-level longitudinal human activity data obtained based on the KU-HAR dataset. It showed that probabilistic models could effectively learn the multi-level variation that was inherent in actual wearable sensor data. In contrast to the classical frequentist, or strictly discriminative models, the Bayesian model enabled a clear separation of uncertainty at global, subject, and activity scales, which offered more statistical understanding of the organization of human movement. It was determined through posterior estimation that inter-subject variation is much higher than inter-activity variation, and hence the significance of personalization in human activity recognition systems. This rank-based understanding comes in handy, especially when it comes to health monitoring software, where personal biomechanics and behavioral characteristics affect sensor dynamics.

The robustness of the model is confirmed by the successful convergence of the core parameters, namely, the global intercept and the residual variance, and the realistic posterior predictive distributions. Although the effects at the activity level were more uncertain, this weakness indicates the intricate overlap of the classes of movement in raw sensor space. However, the posterior predictive tests showed that the model has high generative fidelity, and it can produce exact empirical distributions. This probabilistic interpretability is a clear benefit to black-box machine learning techniques, providing confidence and plausible areas of clinical and decision-making.

In addition to its empirical results, the study supports the methodological argument of hierarchical Bayesian modeling in wearable analytics. The capability of modeling uncertainty, uneven sampling, and incorporating subject-specific effects is in line with the new requirements of adaptive, reliable, and ethically transparent AI in healthcare.

## LIMITATIONS

Future studies need to consider scaling of this method with more sophisticated inference methods and feature-learning priors with Bayesian deep models. To summarize, the paper has solidified Bayesian hierarchy as a statistical method, but more crucially, as a fundamental paradigm on which human activity recognition systems are to be understood and engineered to be deployed in real-life settings, particularly in personal and clinical settings.

## LIST OF ABBREVIATIONS

|             |   |                               |
|-------------|---|-------------------------------|
| <b>CNNs</b> | = | Convolutional Neural Networks |
| <b>ESS</b>  | = | Effective Sample Size         |
| <b>HAR</b>  | = | Human Activity Recognition    |
| <b>HDI</b>  | = | Highest Density Intervals     |
| <b>HMC</b>  | = | Hamiltonian Monte Carlo       |
| <b>IMU</b>  | = | Inertial Measuring Units      |
| <b>MCMC</b> | = | Markov Chain Monte Carlo      |
| <b>NUTS</b> | = | No-U-Turn Sampler             |
| <b>PPC</b>  | = | Posterior Predictive Checks   |
| <b>RNNs</b> | = | Recurrent Neural Networks     |
| <b>SVMs</b> | = | Support Vector Machines       |
| <b>VAEs</b> | = | Variational Autoencoders      |

## AUTHOR'S CONTRIBUTION

U.Y. has contributed to conceptualization of study, development of idea, methodology, analysis of result and interpretation of result.

## AVAILABILITY OF DATA AND MATERIALS

The data used in this study were obtained from the publicly available KU-HAR Human Activity Recognition Dataset, available at: <https://www.kaggle.com/datasets/arashnic/har-1>

## FUNDING

None.

## CONFLICT OF INTEREST

The author declares that there is no conflict of interest regarding the publication of this article.

## ACKNOWLEDGEMENTS

Declared none.

## DECLARATION OF AI

During the preparation of this manuscript, the author used ChatGPT to assist with language editing and refinement. All generated content was subsequently reviewed, revised, and verified by the author, who takes full responsibility for the accuracy, integrity, and final content of the manuscript.

## REFERENCES

- [1] Gates KM, Chow SM, Molenaar PCM. Intensive Longitudinal Analysis of Human Processes. Chapman and Hall/CRC; 2023. <https://doi.org/10.1201/9780429172649>
- [2] Picerno P, Iosa M, D'Souza C, Benedetti MG, Paolucci S, Morone G. Wearable inertial sensors for human movement analysis: a five-year update. *Expert Rev Med Devices*. 2021; 18(sup1): 79-94. <https://doi.org/10.1080/17434440.2021.1988849>
- [3] Mukund A. Wearable technology and sensor data in assistive systems: enhancing rehabilitation through predictive analytics. In: *Predictive Algorithms for Rehabilitation and Assistive Systems*. IGI Global Scientific Publishing; 2025. p. 253-290. <https://doi.org/10.4018/979-8-3373-0194-5.ch010>
- [4] Xu BS, Yang XM, Zou AC, Zang CP. Efficient metamodeling and uncertainty propagation for rotor systems by sparse polynomial chaos expansion. *Probab. Eng. Mech.* 2025; 79: 103723. <https://doi.org/10.1016/j.probengmech.2024.103723>
- [5] Inchausti P. *Statistical Modeling With R: A Dual Frequentist and Bayesian Approach for Life Scientists*. Oxford University Press; 2022. <https://doi.org/10.1093/oso/9780192859013.001.0001>
- [6] Riezler S, Hagmann M. *Validity, Reliability, and Significance: Empirical Methods for NLP and Data Science*. Springer Nature; 2024. <https://doi.org/10.1007/978-3-031-57065-0>
- [7] Xing L, Gardoni P, Zhou Y, Zhang P. DNN-metamodeling and fragility estimate of high-rise buildings with outrigger systems subject to seismic loads. *Reliab. Eng. Syst. Saf.* 2025; 253: 110572. <https://doi.org/10.1016/j.res.2024.110572>

- [8] Mukherjee A. New Bayesian methods for longitudinal data analysis with complex dependence structures. *Electronic Theses and Dissertations*. 2025; 4572. Available from: <https://ir.library.louisville.edu/etd/4572> (Accessed on: 19 November 2025).
- [9] Veenman M, Stefan AM, Haaf JM. Bayesian hierarchical modeling: an introduction and reassessment. *Behav Res Methods*. 2024; 56(5): 4600-4631. <https://doi.org/10.3758/s13428-023-02204-3>
- [10] Marwala T, Mbuva R, Mongwe WT. *Hamiltonian Monte Carlo Methods in Machine Learning*. Elsevier; 2023. <https://doi.org/10.1016/C2021-0-02845-5>
- [11] Azad MS, Nguyen DD, Thusa B, Lee TH. Variance-based sensitivity of seismic damage of the containment building using efficient Bayesian additive regression trees. *Struct Des Tall Spec Build*. 2025; 34(6): e70028. <https://doi.org/10.1002/tal.70028>
- [12] Mukhopadhyay P, Majumdar K, Basu S. Human activity recognition using advanced machine learning and deep learning techniques. In: Das N, Khan AK, Mandal S, Krejcar O, Bhattacharjee D, (eds). *Proceedings of International Conference on Data, Electronics and Computing*. Springer Nature; 2025. p. 155-177. [https://doi.org/10.1007/978-981-97-8476-9\\_12](https://doi.org/10.1007/978-981-97-8476-9_12)
- [13] Mason AJ, Gomes M, Carpenter J, Grieve R. Flexible Bayesian longitudinal models for cost-effectiveness analyses with informative missing data. *Health Econ*. 2021; 30(12): 3138-3158 <https://doi.org/10.1002/hec.4408>
- [14] Mara CA. Methods for analyzing longitudinal data from randomized pretest-posttest-follow-up trials in behavioral research: a practical guide to latent change models. *J Behav Med*. 2025; 49: 286-297. <https://doi.org/10.1007/s10865-025-00600-y>
- [15] Murphy JI, Weaver NE, Hendricks AE. Accessible analysis of longitudinal data with linear mixed effects models. *Dis Model Mech*. 2022; 15(5): dmm048025. <https://doi.org/10.1242/dmm.048025>
- [16] Brysbaert M, Debeer D. How to run linear mixed effects analysis for pairwise comparisons? A tutorial and a proposal for the calculation of standardized effect sizes. *J Cogn*. 2025; 8(1): 5. <https://doi.org/10.5334/joc.409>
- [17] Vakıtbilir N, Islam A, Gomez A, Stein KY, Froese L, Bergmann T, *et al*. Multivariate modelling and prediction of high-frequency sensor-based cerebral physiologic signals: narrative review of machine learning methodologies. *Sensors*. 2024; 24(24): 8148. <https://doi.org/10.3390/s24248148>
- [18] Seedorff N, Brown G, Scorza B, Petersen CA. Joint Bayesian longitudinal models for mixed outcome types and associated model selection techniques. *Comput Stat*. 2023; 38(4): 1735-1769. <https://doi.org/10.1007/s00180-022-01280-x>
- [19] Li Z. A review of Bayesian posterior distribution based on MCMC methods. In: Cao W, Ozcan A, Xie H, Guan B, (eds). *Computing and Data Science*. Springer Nature; 2021. p. 204-213. [https://doi.org/10.1007/978-981-16-8885-0\\_17](https://doi.org/10.1007/978-981-16-8885-0_17)
- [20] Winter S, Campbell T, Lin L, Srivastava S, Dunson DB. Emerging directions in Bayesian computation. *Stat Sci*. 2024; 39(1): 62-89. <https://doi.org/10.1214/23-STS919>
- [21] Bibbò L, Vellasco MMBR. Human activity recognition in healthcare. *Appl Sci*. 2023; 13(24): 13009. <https://doi.org/10.3390/app132413009>
- [22] Moreno MP. Probabilistic models for human behavior learning. 2021. Available from: <https://hdl.handle.net/10016/32935> (Accessed on: 19 November 2025).
- [23] Kulsoom F, Narejo S, Mehmood Z, Chaudhry HN, Butt A, Bashir AK. A review of machine learning-based human activity recognition for diverse applications. *Neural Comput Applic*. 2022; 34(21): 18289-18324. <https://doi.org/10.1007/s00521-022-07665-9>
- [24] Register B. Comparing the effectiveness of standard vs. multilevel machine learning algorithms on hierarchical data. *Theses and Dissertation*. 2025. <https://doi.org/10.13016/mv4w-liso>
- [25] Pitombeira-Neto AR, França DS de, Cruz LA, Silva TLC da, Macedo JFA de. An ensemble Bayesian dynamic linear model for human activity recognition. *IEEE Access*. 2025; 13: 30316-30333. <http://doi.org/10.1109/ACCESS.2025.3541385>
- [26] Ni J, Tang H, Haque ST, Yan Y, Ngu AHH. A survey on multimodal wearable sensor-based human action recognition. *arXiv*. 2024. <https://doi.org/10.48550/arXiv.2404.15349>
- [27] Balaha HM, Hassan AES. Comprehensive machine and deep learning analysis of sensor-based human activity recognition. *Neural Comput Appl*. 2023; 35(17): 12793-12831. <http://doi.org/10.1007/s00521-023-08374-7>
- [28] Zhang S, Li Y, Zhang S, Shahabi F, Xia S, Deng Y, *et al*. Deep learning in human activity recognition with wearable sensors: a review on advances. *Sensors*. 2022; 22(4): 1476. <https://doi.org/10.3390/s22041476>
- [29] Gholamiangonabadi D, Kiselow N, Grolinger K. Deep neural networks for human activity recognition with wearable sensors: leave-one-subject-out cross-validation for model selection. *IEEE Access*. 2020; 8: 133982-133994. <https://doi.org/10.1109/ACCESS.2020.3010715>
- [30] Gong D, Medioni G, Zhao X. Structured time series analysis for human action segmentation and recognition. *IEEE Trans Pattern Anal Mach Intell*. 2014; 36(7): 1414-1427. <https://doi.ieeecomputersociety.org/10.1109/TPAMI.2013.244>
- [31] Yun W, Li F, Pan Y, Zhang H. A sequential metamodel-based importance sampling coupled with adaptive Kriging model method for efficiently estimating the global reliability sensitivity indices. *Probab. Eng. Mech*. 2025; 82: 103848. <https://doi.org/10.1016/j.proengmech.2025.103848>
- [32] Al Amer FM, Thompson CG, Lin L. Bayesian methods for meta-analyses of binary outcomes: implementations, examples, and impact of priors. *Int J Environ Res Public Health*. 2021; 18(7): 3492. <https://doi.org/10.3390/ijerph18073492>

- [33] Griffiths T, Conti ZX, Wilson C, *et al.* Decision support for engineering and design in a fusion pilot-plant concept using Bayesian networks as meta-models. *Nucl Fusion*. 2025; 65(6): 066019. <https://doi.org/10.1088/1741-4326/add549>
- [34] Scheurer S, Tedesco S, O'Flynn B, Brown KN. Comparing person-specific and independent models on subject-dependent and independent human activity recognition performance. *Sensors*. 2020; 20(13): 3647. <https://doi.org/10.3390/s20133647>
- [35] Bendarkar MV, Sarojini D, Mavris DN. Off-nominal performance and reliability of novel aircraft concepts during early design. *J Aircr*. 2022; 59(2): 400-414. <https://doi.org/10.2514/1.C036395>
- [36] Le BV, Nguyen A, Richter O, Nguyen TT. Comparison of frequentist and Bayesian generalized linear models for analyzing the effects of fungicide treatments on the growth and mortality of *Piper nigrum*. *Agronomy*. 2021; 11(12): 2524. <https://doi.org/10.3390/agronomy11122524>
- [37] Shen Y. Understanding contributions of indirect and direct evidence to statistical power in Bayesian network meta-analysis: Simulation studies and real-world applications [Master's Thesis]. Duke University; 2024. Available From: <https://hdl.handle.net/10161/31009>
- [38] Uzun ET, Hızal Ç, Aktaş E. Reliability assessment of structures with Bayesian model updating accelerated *via* polynomial-chaos-kriging metamodeling. *Struct. Infrastruct. Eng.* 2025: 1-22. <https://dx.doi.org/10.1080/15732479.2025.2474115>
- [39] Zhang Y, Chen J. Dual-loop integration framework for model-based system design and reliability analysis using Bayesian networks. *Results Eng.* 2025; 27: 106018. <https://doi.org/10.1016/j.rineng.2025.106018>
- [40] Möbius. Human Activity Recognition Dataset [dataset]. Kaggle; 2022. Available from: <https://www.kaggle.com/datasets/arashnic/har-1>