IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 May 2024, accepted 22 May 2024, date of publication 27 May 2024, date of current version 12 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3405471

RESEARCH ARTICLE

Intelligent Recognition of Multimodal Human Activities for Personal Healthcare

S. R. SANNASI CHAKRAVARTHY[1], N. BHARANIDHARAN[2], V. VINOTH KUMAR[2], (Member, IEEE), T. R. MAHESH[3], (Senior Member, IEEE), SURBHI BHATIA KHAN[4,5], (Senior Member, IEEE), AHLAM ALMUSHARRAF[6], AND EID ALBALAWI[7]

[1]Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam 638401, India
[2]School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore 632014, India
[3]Department of Computer Science and Engineering, JAIN (Deemed-to-be University), Bengaluru 562112, India
[4]School of Science Engineering and Environment, University of Salford, M5 4WT Salford, U.K.
[5]Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon
[6]Department of Management, College of Business Administration, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[7]College of Computer Science and Information Technology, King Faisal University, AlAhsa 400-31982, Saudi Arabia

Corresponding authors: Surbhi Bhatia Khan (s.khan138@salford.ac.uk) and Ahlam Almusharraf (aialmusharraf@pnu.edu.sa)

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R432), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Nowadays, the advancements of wearable consumer devices have become a predominant role in healthcare gadgets. There is always a demand to obtain robust recognition of heterogeneous human activities in complicated IoT environments. The knowledge attained using these recognition models will be then combined with healthcare applications. In this way, the paper proposed a novel deep learning framework to recognize heterogeneous human activities using multimodal sensor data. The proposed framework is composed of four phases: employing dataset and processing, implementation of deep learning model, performance analysis, and application development. The paper utilized the recent KU-HAR database with eighteen different activities of 90 individuals. After preprocessing, the hybrid model integrating Extreme Learning Machine (ELM) and Gated Recurrent Unit (GRU) architecture is used. An attention mechanism is then included for further enhancing the robustness of human activity recognition in the IoT environment. Finally, the performance of the proposed model is evaluated and comparatively analyzed with conventional CNN, LSTM, GRU, ELM, Transformer and Ensemble algorithms. To the end, an application is developed using the Qt framework which can be deployed on any consumer device. In this way, the research sheds light on monitoring the activities of critical patients by healthcare professionals remotely. The proposed ELM-GRUaM model achieved supreme performance in recognizing multimodal human activities with an overall accuracy of 96.71% as compared with existing models.

INDEX TERMS Artificial intelligence, consumer electronics, deep learning, healthcare, human activity recognition, IoT, multimodal data.

I. INTRODUCTION

Intelligent Decision Support Systems (IDSS) provide effective solutions to many of the challenges currently being faced around the world. The widespread adoption of machine learning and deep learning techniques has greatly facilitated the development of IDSS, due to the easier availability of multiple datasets related to various aspects of human lives [1]. These developments have proven to be more valuable during the fight against the COVID-19 crisis. IDSS plays a crucial role in helping healthcare professionals to detect diseases earlier, thereby increasing the chances of patient survival.

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

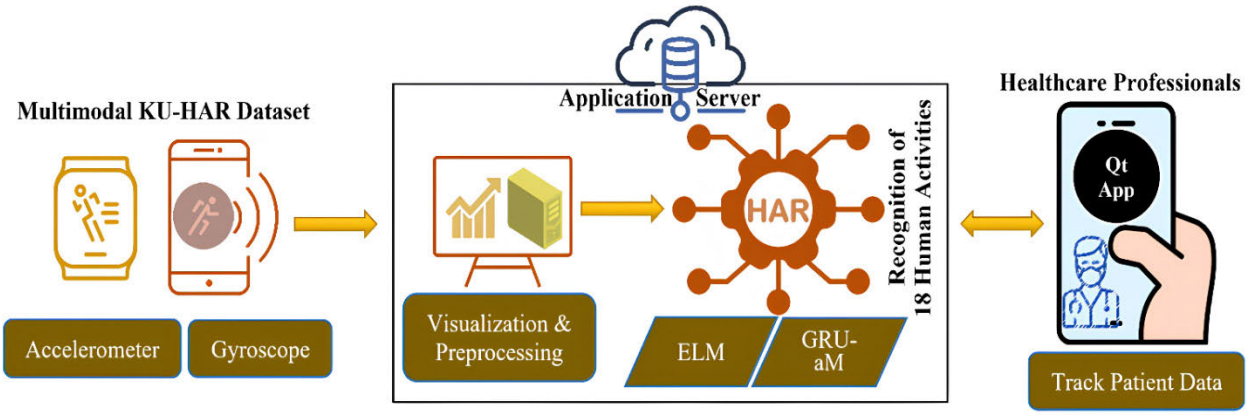

FIGURE 1. Usage of multimodal patient data for activity monitoring in an IoT environment.

In addition, the IDSS can be utilized in patients' gesture recognition and has been distinguished in smart healthcare, ensuring timely response to patient needs especially remote control of resources [2]. In under-developed nations with poor or inadequate health services, IDSS is an important solution to provide affordable and cost-effective services that do not require expensive equipment or trained personnel.

In today's scenario, there is an unusually large number of IoT-enabled devices aimed at improving decision-making processes in complex systems. Rapid miniaturization and development of sensors, reduced energy requirements, revolutionized the Human Activity Recognition (HAR) field in the detection of vulnerable diseases such as diabetes [3] and heart disease [4] early, and the first symptoms of COVID-19 using data from sensors [5] on smartwatches. Moreover, these advancements are becoming real and implemented to assist several healthcare requirements.

Recently, there have been tremendous advancements in the field of medicine involves transitioning away from the one-size-fits-all approach to embrace Personalized Health Care and medicine [6], [7]. This advancement has been driven by the aging population and the escalating costs associated with chronic diseases. Accordingly, a promising solution is always in demand to address these challenges. The required solution necessitates innovative methodologies for continuously monitoring and assessing the vital signs of individual patients, allowing for the customization of medication plans tailored to specific needs. This can be attained through the utilization of Machine Learning, Deep Learning, and the Internet of Things approaches by deploying appropriate sensors around the patient. These sensors would continuously transmit data to healthcare professionals and hospitals, enabling them to make well-informed decisions. This invaluable information serves to empower patients to manage their daily activities more effectively [8]. In this way, the work focuses on the solution for intelligent recognition of multimodal human activities for personal healthcare as illustrated in Figure 1. The key contributions of the proposed work are given below:

(i) For creating a robust and discriminative representation of the input sensor data, the ELM used for feature transformation is integrated with the Gated Recurrent Unit (GRU) used for sequence modeling.

(ii) The integration of GRU allows the framework in capturing complex temporal patterns corresponding to human activity recognition.

(iii) Furthermore, the output predictions are improved by incorporating the attention mechanism (ELM-GRUaM) that provides enhanced focus on relevant parts of the input.

(iv) Thus the combined ELM-GRUaM framework provides effective HAR task recognition in diversified environments and across different individuals.

II. RELATED WORKS

Several researchers have attempted to recognize distinct human activities through different approaches. Some of them are discussed in this section. The authors [9] utilized a deep recurrent neural network (DRNN) architecture for recognizing human activities. The architecture is employed to capture the complete longer-range of input relationships instead of being limited to the kernel window size. Moreover, the work utilized DRNN in cascading, bidirectional, and unidirectional fashions. The UCI-HAR database was utilized and provided a maximum accuracy of 96% as compared with Support Vector Machines (SVM). The researchers [10] introduced an ensemble approach of combining Naïve Bayes, Decision Tree, and kNN models. The approach was implemented using heuristic-based hand-crafted features taken from gyro, magnetometer, and accelerometer devices. The implemented ensemble model is found to be highly sensitive to data collected through distinct people, overlapping, and window size.

The researchers [11] introduced a modern hybridized evolutionary algorithm that combines Genetic Algorithm (GA) with effective evolutionary methods. An implementation of a decision support system was done to help clinicians to manage regular activities. Through a comprehensive empirical study, they demonstrated the effectiveness of their approach for solving models related to healthcare intelligent solutions. The authors [12] proposed a human activity recognition (HAR) model based on an LSTM framework, aimed at improving assistance in the IoT environment. A grid was

added, solving the challenge of limited labeled data. Mixing the user's body sensor data with environmental data, their model achieved the best results compared to other methods such as Random Forest (RF), DNN, and SVM.

The authors [13] developed a dedicated HAR model with the inputs taken from different sensor devices. The model alters the collected time-series inputs from sensor devices as image inputs. These converted image inputs are adopted for maintaining the necessary pattern and feature vectors for solving the problem. For training and evaluation, the researchers utilized a fusion residual architecture through the integration of dual network models. The research yielded better performance with 93% and 98% accuracies on the HAR and MHEALTH datasets. The researchers [14] followed a methodology for HAR tasks through non-invasive means of collecting human movements using video frames. The implementation is done in the following ways: The initial phase is an offline approach for generating binary CNN architectures for recognition. The next is the inference phase which deals with human recognition and their movements by using CNNs. In this way, the research outpaced other approaches with 56% accuracy on the UCF-ARG database.

The researchers [15] employed neural network models to approximate the time-dependent distributions within non-Markovian models. They achieved this by utilizing solutions from simpler, time-inhomogeneous Markovian models, a process that preserves model dimensionality while enabling the inference of kinetic parameters. This neural network was trained using a limited set of noisy measurements obtained from either experimental data or stochastic simulations of the non-Markovian model. As a result, their findings demonstrated that the neural network successfully learned Markovian models capable of accurately representing stochastic dynamics across a spectrum of models. The authors [16] introduced AHAR (Adaptive Human Activity Recognition), energy-efficient CNNs optimized for least-power edge devices. AHAR employs an adaptive design during the inference phase, allowing it to select specific components from its baseline architecture intelligently. The model is evaluated on two databases. For the Opportunity dataset, AHAR achieved better weighted F1 scores of 91.7%, whereas it achieved a supreme 91.5% for the w-HAR dataset.

The authors [17] developed a hybrid model based on the combination of CNN and LSTM for solving HAR tasks. This is applied to a database that integrates the activity samples of 20 people with twelve distinct classes. The study revealed that it provided a maximum performance of 90.8% accuracy over others. The key findings of the current state-of-the-art works related to human activity recognition are comparatively summarized in Table 1. In this way, the paper proposed a deep learning-based architecture integrating the power of ELM and GRU with an attention mechanism to recognize 18 distinct human activities using multimodal data as given in Figure 1.

TABLE 1. Comparative summary of the related works.

Study	Approach	Key Features
[9]	Deep Recurrent Neural Network (DRNN) architecture with UCI-HAR data	Maximum accuracy of 96% compared to Support Vector Machines (SVM)
[10]	Ensemble approach combining Naïve Bayes, Decision Tree, and kNN models	Highly sensitive to data collected through distinct people, overlapping, and window size
[11]	Hybridized evolutionary algorithm combining Genetic Algorithm (GA) with effective evolutionary methods with Healthcare data	Effectiveness demonstrated in solving models related to healthcare intelligent solutions
[12]	Long Short-Term Memory (LSTM) framework with grid addition in IoT environment	Achieved best results compared to other methods such as Random Forest (RF), DNN, and SVM
[13]	Model converting time-series inputs from sensor devices into image inputs, utilizing fusion residual architecture with HAR and MHEALTH data	Better performance with 93% and 98% accuracies on HAR and MHEALTH datasets
[14]	CNN architectures for recognition of human movements from video frames with UCF-ARG data	Outpaced other approaches with 56% accuracy on the UCF-ARG database
[15]	Neural network models approximating time-dependent distributions within non-Markovian models with experimental data	Successful learning of Markovian models accurately representing stochastic dynamics
[16]	Adaptive Human Activity Recognition (AHAR), energy-efficient CNNs optimized for least-power edge devices	Achieved better weighted F1 scores of 91.7% and 91.5% on the Opportunity and w-HAR datasets
[17]	Hybrid model based on a combination of CNN and LSTM for solving HAR tasks on a database with activity samples of 20 people	Provided maximum performance of 90.8% accuracy over others

III. MATERIALS AND METHODS

A. KU-HAR – INPUT DATASET

KU-HAR dataset [18] is an open-source database containing eighteen distinct heterogeneous activity information acquired from nighty people with a mixture of genders. The dataset comprises multimodal data acquired through Gyroscope and accelerometer sensors available in smartphones. It consists of 1,945 rawly collected activity sample values and 20,750 sub-samples extricated from the involved people. Each of these data has three seconds of non-overlapping information about the respective activity [18]. The output classes correspond to 18 distinct human activities are sitting, standing, talking with hand movements while standing or walking, talking with hand movements while sitting, performing sit-ups,

performing full push-ups, repeatedly lying down and standing up, laying still, repeatedly sitting down and standing up, running 20 meters, walking along a circular path, walking 20 meters, walking backward for 20 meters, jumping repeatedly, picking up an object from the floor, playing table tennis, descending from a set of stairs, and ascending on a set of stairs.

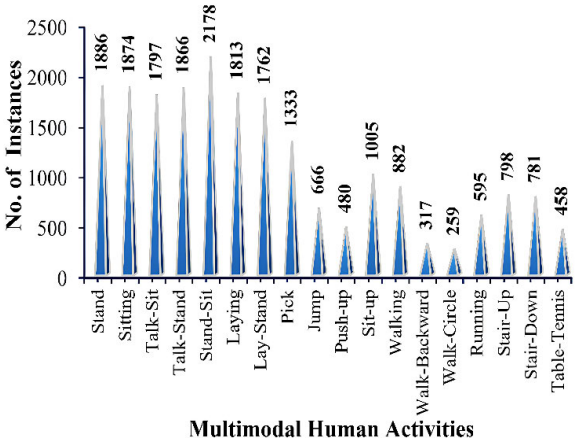

FIGURE 2. Distribution of KU-HAR dataset.

B. VISUALIZATION AND PREPROCESSING OF KU-HAR DATA

Data visualization is much more useful in both analysis and solving real-world problems. The paper visualizes the KU-HAR dataset for predicting its nature. The distribution of input data is illustrated in Figure 2. This plot reveals that the input consists of 18 distinct classes (y-axis) with a total of 20750 samples (x-axis). And the volume of data to be handled is larger and the multimodal data composition of the KU-HAR dataset is plotted in Figure 3. This plot illustrates that columns 1 to 900 represent the sensor readings from the accelerometer axes (X, Y, Z) and 901 to 1800 has the sensor readings from the gyroscope axes (X, Y, Z). A sample visualization of multimodal readings of the KU-HAR dataset is plotted in Figure 4. Here, the plot reveals how the accelerometer (acceleration) and gyroscope data (angular rotation) change over time during the ''Jump'' activity for the 20001st sample. The two plots of Figure 4 illustrate the data from different sensor axes separately and give the idea of understanding the motion characteristics of the activity. This illustration portrays that the data collected from the two sensors with different axes have a nature of highly overlapping. Thus, the paper involves in employing the Extreme Learning Machine (ELM) and Gated Recurrent Unit (GRU) with an Attention Mechanism (aM) for capturing the input patterns efficiently and hence making a robust recognition of multimodal human activities.

The strip plot visualization of sample data is plotted in Figure 5 which provides a visualization of the distribution of sensor data across various activities. For the huge volume of collected overlapping data with different scales as

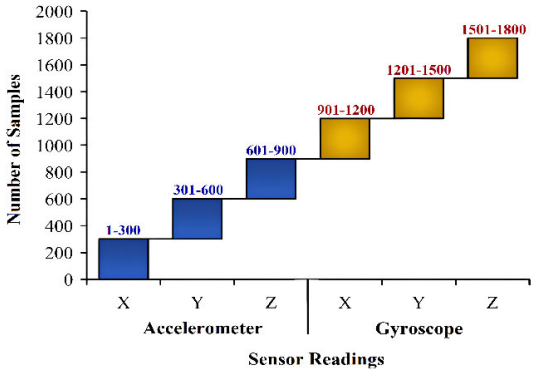

FIGURE 3. Multimodal composition of input dataset.

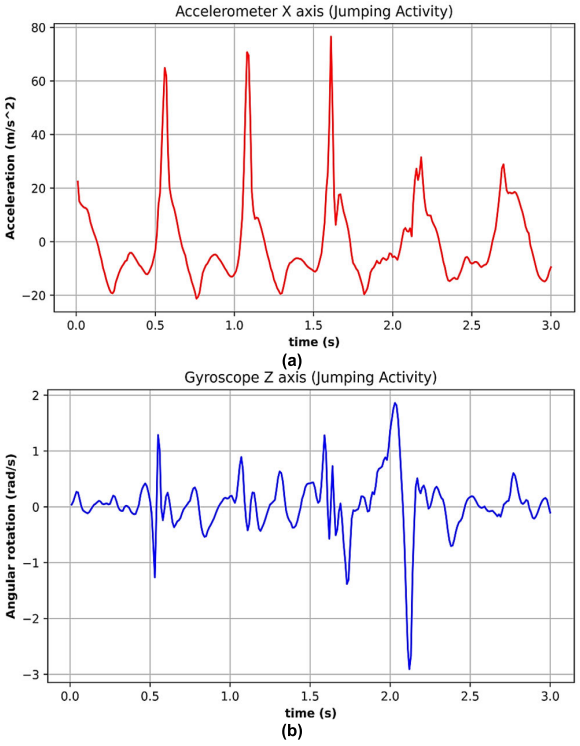

FIGURE 4. (a) Sample data from accelerometer X axis (b) Sample data from gyroscope Z axis representing jumping activity of human.

shown in Figure 5, the standardization of data is mandatory for further classification process. The paper employed the StandardScaler [19] function of the sklearn library which transforms the data through the removal of mean and scaling to unit variance. The standardization [19] is performed using the mathematical operation as presented in Equation (1).

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

In Equation (1), x is the data inputs, μ and σ represent the mean and standard deviation. In addition, the dataset is checked for any missing values but there are no missing values found. In addition, gravity acceleration data were ignored from the accelerometer information, so no filters were employed to get rid of noise [18].

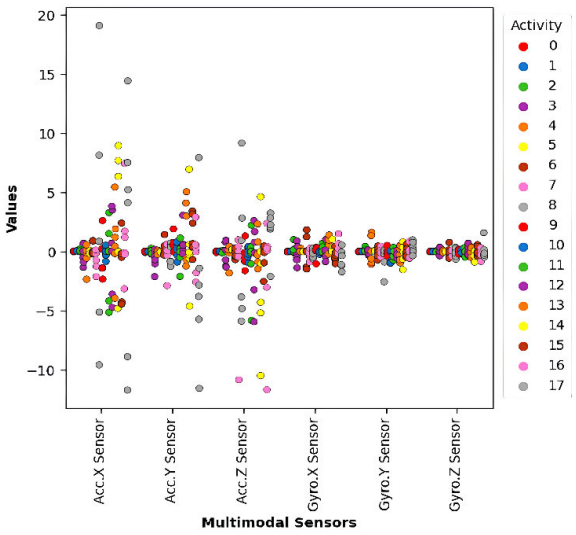

FIGURE 5. Sample data visualization of multimodal dataset.

IV. PROPOSED METHOD

The proposed Extreme Learning Machine-Gated Recurrent Unit with Attention Mechanism (ELM-GRUaM) model used for multimodal human activity recognition will be explained here with different sub-sections.

A. NEED FOR FEATURE TRANSFORMATION USING EXTREME LEARNING MACHINE (ELM)

The multimodal data from the gyroscope and accelerometer sensors are found to be highly complex (as illustrated in the plot of Figure 4) and non-linear (as shown in the plot of Figure 5). In this research, ELM introduces non-linearity through its hidden layer activation function (ReLU). This makes ELM to capture complex patterns and relationships in the data that might not be captured effectively by linear transformations. This enhanced representation supports improved recognition of human activities and thus making it easier for subsequent models to classify activities accurately.

B. EXTREME LEARNING MACHINE (ELM) FOR FEATURE TRANSFORMATION

The architecture of ELM used for feature transformation is given in Figure 6. As in the plot, ELM's architecture resembles the neural network models but ELM is faster than the traditional neural network models [20]. The mathematical background of ELM used in this paper for feature transformation will be discussed next. For an input feature vector, $x \in \mathbb{R}^n$ with n features, a hidden layer with m neurons, and an output layer having p neurons, the ELM model computes the output, $y \in \mathbb{R}^p$ as given below.

As illustrated in Equation (2), for each neuron i in the hidden layer [21],

$$z_i = \sum_{j=1}^{n} w_{ij} x_j + b_i \tag{2}$$

In Equation (2), w_{ij} and b_i represent the weight connecting feature j to the hidden neuron i and bias of the hidden neuron i.

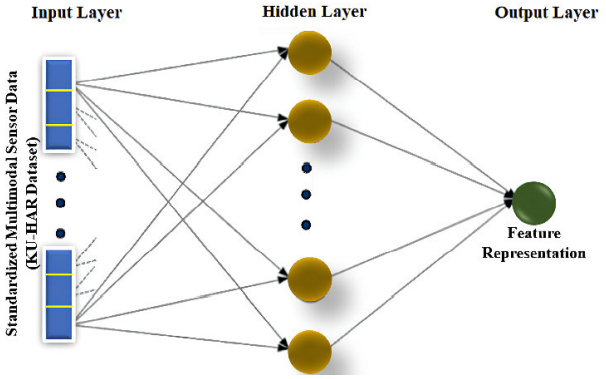

FIGURE 6. ELM architecture for feature transformation.

z_i refers to the net input to hidden neuron i. In addition, the paper employed the activation function used in the hidden layer as the ReLU (Rectified Linear Unit) function as presented in Equation (3).

$$h_i = ReLU(z_i) \tag{3}$$

The weights connecting the input to the hidden layer are fixed and randomly generated but the weights connecting the hidden layer to the output layer will be learned through training. Finally, the output of the ELM model is calculated as a weighted sum of the activations of the hidden layer neurons as presented in Equation (4):

$$y_k = \sum_{i=1}^{m} w_{ki}^{out} h_i + b_k^{out} \tag{4}$$

In Equation (4), w_{ki}^{out} refers to the weight connecting the hidden neuron i to the output neuron k. b_k^{out} denotes the bias of the output neuron k. And y_k indicates the final computation of the output neuron k. It is to be noted that the ELM used here is not for classification tasks but for feature transformation. That is, ELM simply transforms the input data into a newer feature space characterized by the activations h_i of the hidden layer neurons. These activations must serve as the transformed features. Now these are fed as input to the subsequent GRU network with an attention mechanism for multimodal human activity recognition. As a result, the above ELM-transformed features will support the research to capture relevant information and patterns from the multimodal sensor data.

C. ELM-GATE RECURRENT UNIT (GRU)

The proposed framework for multimodal human activity is illustrated in Figure 7. As in Figure, the ELM transformed features are fed as input to the GRU network. Human activities are basically temporal and sequential. The readings were recorded over time, and capturing the dependencies and patterns within sequences of multimodal sensor data is necessary for improved recognition. The paper chooses GRU as the recurrent neural network (RNN) model since it is good at modeling sequential data [22]. In addition, GRU involves fewer parameters as compared to LSTM architecture. This

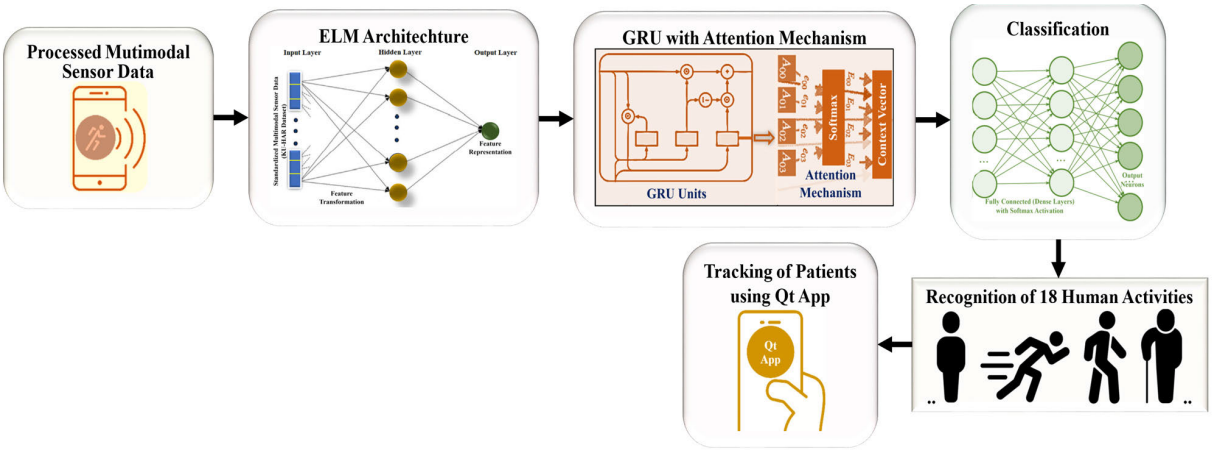

FIGURE 7. Proposed framework for multimodal human activity recognition.

makes GRU as computationally efficient and effective at capturing long-term dependencies [23].

As in Figure 7, GRU starts with an initial hidden state (h_0) and this will be initialized either to zeros or learned as a parameter of the model. The update (z_t) and reset gates (r_t) control how the data is conceded between time-steps and this makes GRU to selectively reset and update its hidden states. Here, z_t helps in deciding how much of the previous hidden states h_{t-1} are to be retained and how much of the new candidate hidden states $\tilde{h}_t$ are to be added to the current state.

This can be mathematically illustrated as shown in Equation (5),

$$z_t = \sigma(W_z.[h_{t-1}, x_t] + b_z) \tag{5}$$

In Equation (5), the sigmoid activation function is denoted as σ, W_z, and b_z represent learned weights and biases. And $[h_{t-1}, x_t]$ indicates the concatenation of the previous hidden state (h_{t-1}) and the current input (x_t). Next, r_t helps in deciding how much of the previous hidden states h_{t-1} are to be forgotten during the computation of $\tilde{h}_t$. This can be mathematically illustrated as given in Equation (6).

$$r_t = \sigma(W_r.[h_{t-1}, x_t] + b_r) \tag{6}$$

The value of $\tilde{h}_t$ can be computed based on r_t and x_t. This can be mathematically illustrated as given in Equation (7).

$$\tilde{h}_t = \tanh(W_h.[r_t \odot h_{t-1}, x_t] + b_h) \tag{7}$$

In Equation (7), *tanh* denotes the hyperbolic tangent activation function. The final hidden state h_t can be computed as illustrated in Equation (8).

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{8}$$

The GRU processes the entire sequence one-time step at a time and so updating its hidden state h_t at each step. This ensures the capturing of temporal dependencies with h_t encoding information about the sequence upto the current-time step.

D. ELM- GRU WITH ATTENTION MECHANISM (aM)

Now, the attention mechanism considers the GRU's sequence of hidden states as its input. For each time-step, t, an attention score (e_t) is calculated. This score quantifies the relevance of the t^{th} hidden state to the current context. This could be done by comparing each hidden state with a context vector from the previous time-step. The e_t scores are often passed via a softmax function to determine attention weights (a_t) ensuring that they sum to one. Then, the context vector represented as c_t can be determined as a weighted addition of the GRU hidden states; mathematically illustrated as given in Equation (9).

$$c_t = \sum_{i=1}^{T} a_i.h_i \tag{9}$$

In Equation (9), the total number of time steps in the sequence is denoted as T and the attention weights are represented as a_i. The c_t from the attention mechanism captures the most relevant information from the GRU hidden states for the current context. The c_t is conceded to the next time-step, influencing the attention scoring for the subsequent hidden state, making the mechanism recurrent.

Finally, the classification layer as shown in Figure 7 is responsible for producing a probability distribution over the activity classes. The layer is composed of fully connected (dense) layers that can further process the context vector and extract relevant features for classification. The final layer is a typical softmax function for multi-class classification. This function determines the probability distribution over all possible targets based on the input context vector. Each output neuron corresponds to one activity target, and the softmax function normalizes the values across these neurons for ensuring that they sum to one. The steps involved in the overall proposed framework as discussed above are illustrated in Algorithm 1.

V. RESULTS AND DISCUSSION

The outcomes of the recognition of multimodal human activities using the proposed framework are discussed here. The

Algorithm 1 Overall Proposed Framework for HAR

Input: Multimodal KU-HAR Data
STEP 1: Data Visualization and Preprocessing
Perform data visualization and normalization using Equation (1).
STEP 2: Data Partitioning:
Split data into training and testing sets: X_train, X_test, y_train, y_test with $test\_size = 0.3$.
STEP 3: Creation of ELM Model:
Input layer with Input features ($X\_trainshape$); hidden layer with 150 hidden neurons and ReLU activation as presented in Equations (2) and (3).
STEP 4: Training of ELM Model:
Initialize hidden layer weights (w) randomly and learn output layer weights using Equation (4).
STEP 5: Transform Data:
ELM-transformed training and testing data: X_train_elm = ReLU($X\_train \times w + b$) and X_test_elm = ReLU($X\_test \times w + b$).
STEP 6: Give transformed data to GRU Model:
GRU_output = GRU(X_elm) using Equations (5) to (8).
STEP 7: Use attention mechanism:
Calculate attention scores (α) based on GRU outputs (h) and context (c): $e = tanh(W1 \times h + W2 * c)$ and $\alpha = softmax(e)$. Calculate the weighted sum of GRU outputs with attention using Equation (9).
STEP 8: Pass it to output layer:
This layer produces class probabilities using a softmax activation as:
$probabilities = softmax(output\_layer\_input)$.
STEP 9: Tracking of activities using Qt application.

execution of the research work is implemented using Python 3.6 installed on a Windows system with 16-GB of RAM and an Intel Core i7 processor. For comparative analysis, the paper adopted the four existing algorithms namely standalone CNN [24], LSTM [25], GRU [26], ELM [27], Transformer model [35], and Random Forest [18] models. For the phase of classification, the paper employed the stratified partition where a ratio of 70:30 throughout the implementation for making training and testing sets of inputs.

A. PERFORMANCE METRICS AND HYPERPARAMETERS

The paper adopted the performance measures which are derived using the confusion matrix (CM). This matrix comprehends four parameters namely true negative and positives, false negative and positives, characterized as TN, TP, FN, and FP. The metrics adopted are precision, accuracy, and recall [28]. For attaining a better balance between sensitivity and precision, the paper employed an additional metric of F1 score [29] for the experimentations. Finally, the paper utilized Cohen's kappa metric (κ) [30] for further analysis and validation. The number of hidden neurons selection is crucial when ELM is employed for solving time-series problems. Based on the experimental study, the research employed an optimal count of 150 as the number of hidden neurons. The employed GRU model is implemented with 128 units and is set to return sequences. This makes GRU to capture temporal dependencies in the data. The GRU model is compiled with the Adam optimizer and categorical cross-entropy loss function. GRU is trained using the ELM-transformed training data and one-hot encoded labels for ten epochs with a batch-size of 64. An Attention layer is implemented then for weighing the significance of distinct time-steps in the sequence. The attention output is finally concatenated with the GRU output for obtaining final predictions.

B. EXPERIMENTAL RESULTS AND COMPARATIVE ANALYSIS

At an initial point of experimentation, an ablation study is performed against each component as illustrated in Table 2. Here, the combination of ELM+GRU+aM provides the maximum classification performance for the employed problem. The overall cross-validated results using the proposed and existing classification models for the employed problem of multimodal human activity recognition are tabulated in Table 3. One of the ensemble models, the Random Forest (RF) algorithm provided an overall accuracy of 89.64% with a kappa (κ) value of 0.888. The standalone ELM model used for classification provided an overall accuracy of 90.02% with the kappa (κ) value of 0.892. A simple CNN is then employed for classification which gives 90.58% of overall accuracy and 0.899 as kappa. It is to be noted that the performance of all three above models provided overlapping performance. The LSTM, Transformer, and GRU models are then employed where 91.93%, 95.18%, and 93.16% are obtained as overall classification accuracies. As compared with LSTM and Transformer models, GRU performed well for the employed

TABLE 2. Ablation study for multimodal human activity recognition.

Modules	Test_1	Test_2	Test_3	Test_4	Test_5
ELM	✗	✗	✓	✓	✓
CNN	✓	✗	✗	✗	✗
LSTM	✗	✓	✗	✗	✗
GRU	✗	✗	✗	✓	✓
aM with a classification layer	✗	✗	✗	✗	✓
Accuracy (%)	90.58	91.93	90.02	92.44	96.71

TABLE 3. Classifiers' performance for multimodal human activity recognition.

Classifiers	Overall Accuracy (%)	Kappa (κ)
RF	89.64	0.888
ELM	90.02	0.892
CNN	90.58	0.899
LSTM	91.93	0.913
Transformer Model	95.18	0.948
GRU	93.16	0.926
ELM-GRUaM	96.71	0.965

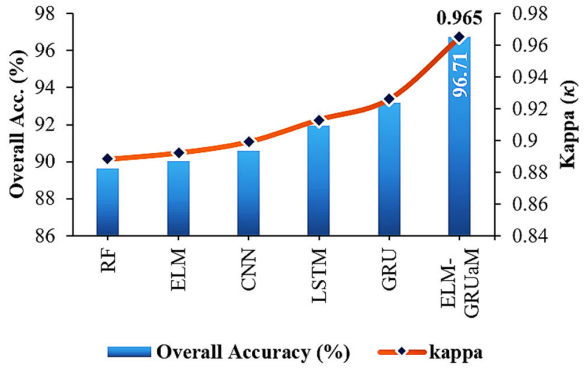

FIGURE 8. Comparative analysis of the proposed model.

human activity recognition and this is due to the nature of computationally efficiency and being effective at capturing long-term dependencies of applied inputs. Thus, the GRU model outperforms the performance of all other models with a validated kappa value of 0.926.

The graphical illustration of comparative performance analysis with validation is plotted in Fig. 8. As in the plot, the proposed architecture of ELM-GRUaM provides a supreme overall classification accuracy of 96.71%. The right choice of integrating ELM for feature representation and GRU for efficient capturing of input patterns provided a better performance than others. Furthermore, the attention mechanism is concatenated with the above model to focus on different parts of the input sequence dynamically. This makes the proposed model, ELM-GRUaM provide a supreme validated kappa value of 0.965. This superior κ value shows that the proposed architecture, ELM-GRUaM gives a perfect agreement for the employed multiclass classification problem as reported in the ablation study.

C. INSIGHT PERFORMANCE ANALYSIS

The discussion made in the previous sub-section is only about the analysis of overall performance for the employed problem. The research involves the classification of 18 distinct human activities collected from multimodal sensor devices. Hence, the individual or insight performance of each classification algorithm needs to be examined for individual output classes, respectively. Furthermore, due to the unavoidable class imbalance as shown in Fig. 2, the insight performance analysis of classification models is essential. This gives us more clarification on where the algorithm's performance lacks or is better in recognizing multimodal human activities. The individual performance analysis of GRU and the proposed ELM-GRUaM models are listed in Tables 4 and 5. Here, the number of classified outputs indicates the number of activities classified as belonging to that class from all the classes. The individual performance comparison of GRU and the proposed ELM-GRUaM model is graphically plotted in Fig. 9.

As from Fig. 9 and Tables 4 and 5, the conventional GRU algorithm performs well for static activities but lacks in recognizing dynamic activities. This problem is tackled using the inclusion of an attention mechanism with the GRU model and thus the proposed ELM-GRUaM model performs well in recognizing multimodal human activities effectively.

TABLE 4. Insight performance analysis of GRU model.

Output Class	Number of Inputs	Number of Classified Outputs	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Standing	566	568	99.20	95.42	95.76	96.11
Sitting	562	559	99.24	96.06	95.55	96.42
Talk-Sit	539	532	99.15	95.67	94.43	95.28
Talk-Stand	560	564	99.26	95.56	96.25	96.19
Stand-Sitting	653	652	99.28	96.62	96.47	97.44
Laying	544	546	99.23	95.42	95.77	96.00
Lay-Stand	529	526	99.15	95.24	94.70	95.17
Pick	400	397	99.18	93.95	93.25	94.23
Jump	200	199	99.34	89.95	89.50	90.36
Push-up	144	153	99.31	83.00	88.19	86.27
Sit-up	301	309	99.42	92.88	95.34	94.42
Walking	265	268	99.34	91.79	92.83	92.26
Walk-Backward	95	99	99.20	72.72	75.78	74.34
Walk-Circle	78	83	99.24	68.67	73.07	71.19
Running	178	184	99.33	86.95	89.88	88.46
Stair-Up	239	237	99.16	89.45	88.70	89.25
Stair-Down	234	230	99.16	89.56	88.03	89.41
Table-Tennis	137	118	99.12	84.74	72.99	78.06

TABLE 5. Insight performance analysis of the proposed ELM-GRUaM model.

Output Class	Number of Inputs	Number of Classified Outputs	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Standing	566	568	99.55	97.35	97.70	98.19
Sitting	562	563	99.57	97.51	97.68	98.47
Talk-Sit	539	541	99.61	97.59	97.96	98.68
Talk-Stand	560	562	99.61	97.68	98.04	98.55
Stand-Sitting	653	652	99.63	98.31	98.86	98.19
Laying	544	542	99.55	97.60	97.74	97.82
Lay-Stand	529	529	99.61	97.73	97.91	98.64
Pick	400	400	99.55	96.50	96.86	96.78
Jump	200	202	99.61	93.56	94.70	94.27
Push-up	144	143	99.53	90.21	89.68	90.81
Sit-up	301	301	99.65	96.34	96.85	96.66
Walking	265	266	99.57	94.73	95.49	95.19
Walk-Backward	95	96	99.53	84.37	85.76	85.16
Walk-Circle	78	85	99.60	81.17	88.47	85.87
Running	178	179	99.60	92.73	93.29	93.17
Stair-Up	239	233	99.61	96.13	93.74	95.55
Stair-Down	234	232	99.55	94.39	93.62	94.13
Table-Tennis	137	130	99.53	91.53	86.88	89.94

D. COMPARISON WITH STATE-OF-THE-ART MODELS FOR HAR

As a final point, the proposed approach is compared against the recently published frameworks and it is listed in Table 6. That is, the recent studies that are associated with the

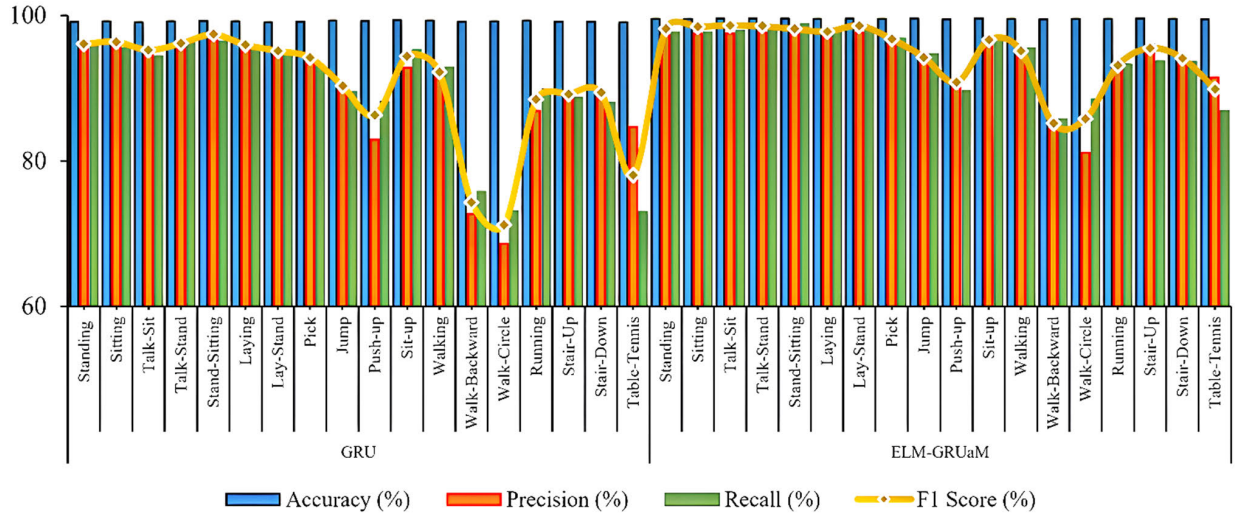

FIGURE 9. Individual performance analysis of GRU and proposed ELM-GRUaM method.

employed research problem are chosen for comparison. Specifically, the works employed with the KU-HAR dataset are taken for comparison. The summary of Table 6 illustrates that the proposed model outperforms the existing works in the employed problem.

TABLE 6. Comparison of the proposed method with the KU-HAR data in recent studies.

Research Works	Methodology	Overall Accuracy (%)
Mahmudul et al. [31]	Handcrafted Feature Extraction with RF Model	89.5
Niloy et al. [28]	Ensemble Learning Algorithm	90
Mahmudul et al. [32]	Feature Prioritization using Wavelets	94.76
Prabhat et al. [33]	Deep-Transfer Heterogenous Model	94.25
Sakorn et al. [34]	ResNet Deep Learning Model	93.54
Proposed Model	ELM-GRUaM	96.71

E. QT APPLICATION FOR HAR

A Qt application is made to deploy applications on any platform such as Windows, Linux, Mac, and Android. The paper employed the Qt framework for developing applications because the applications developed can be deployed across multiple platforms of IoT environments. The front end is designed using QML with a simple user interface. By using this application, healthcare professionals can search the patients either by ID or Name stored from the server for real-time tracking of patients in IoT environments. In addition, hospital practitioners can upload the sensor data for real-time data processing and to track the current activity of patients. The backend implementation is fully taken care of using Python scripts.

F. COMPUTATIONAL COMPLEXITY

The computational complexity of the proposed methodology is given below with the summation of individual component complexity.

ELM: The computational complexity of training an ELM can be $O(MN + MD)$ where the numbers of input neurons, hidden neurons, and training samples are represented by N, M, and D.

GRUaM: The computational complexity of training a GRU layer can be $O(LNT)$ where its number of units, layers, and its input sequence length are represented by N, L, and T. After including K attention heads in attention mechanism, the computational complexity can be changed to $O(KNT)$.

Combining ELM with GRUaM: The combination involves training both an ELM and a GRUaM sequentially. So the overall computational complexity can be computed as $C_{ELM} + C_{GRUaM}$.

G. DISCUSSION OF THE FINDINGS

The summary of the research intended to improve the performance of GRU for robust recognition of multimodal human activities.

- The research novelty lies in enhancing the classification performance through the successful implementation of the proposed ELM-GRU with Attention Mechanism (ELM-GRUaM).
- The above framework is then utilized to improve the s-GRU's classification performance in predicting eighteen distinct human activities collected through multimodal sensor devices.
- As compared with recent existing methods, the proposed algorithm performed well for the employed multiclass classification problem.
- The study compared the classification performance of the proposed framework with conventional standalone models and revealed that the ELM-GRUaM outperforms them, and thereby establishing the novelty of the framework.

H. LIMITATIONS OF THE PROPOSED WORK

Many real-time problems always demand more accurate outcomes either using deep learning or machine learning

architectures. Researchers all around the world are working towards providing promising solutions. Accordingly, the proposed methodology is implemented successfully and robust outcomes have been attained. However, from the plot of Figure 9, it is inferred that the performance of the proposed framework lags at recognizing dynamic activities such as playing as compared with static activities. This reveals that the proposed framework requires to be enhanced further for correctly recognizing dynamic activities (playing, walking) minimizing the risk of recognizing dynamic as static activities. In addition, the computational complexity of the proposed methodology as discussed in Sub-Section F will be minimized without compensating the performance. Subsequently, these problems will be taken into consideration in our future research.

VI. CONCLUSION AND FUTURE WORK

The research proposed a novel deep-learning-based robust framework for recognizing multimodal human activities. For evaluation, the research utilized the KU-HAR dataset which comprises of multimodal sensor (accelerometer and gyroscope) data. The normalized and preprocessed data are initially subjected to feature transformation using Extreme Learning Machine (ELM) model. The obtained feature representation supports to capture of significant patterns and characteristics in the sensor data. This can aid in the HAR task, as they have undergone a non-linear transformation. The ELM-transformed features are then applied to Gated Recurrent Units (GRU). The GRU intakes these features and leverages its sequential modeling capabilities to capture temporal dependencies and recognize human activities over time efficiently. In addition, an Attention Mechanism is concatenated with GRU for assigning distinct weights to each time step's output, representing the significance of each time step's contribution to the final classification decision. In this way, the proposed ELM-GRUaM model provided a supreme outcome of 96.71% as overall classification accuracy with a validating kappa score of 0.965. Furthermore, the robustness of the proposed framework is analysed using insight performance and comparative analysis. A Qt application is developed with QML as frontend and Python scripts as backend for real-time tracking of patients by healthcare professionals.

The future extension of the work will be in the direction of providing a better user interface with additional security for the developed application. The computational complexity of the proposed methodology will be reduced without compensating the performance. Furthermore, the study suggests future researchers to utilize the variants of crow-search algorithm for feature selection since the dataset is larger.

ACKNOWLEDGMENT

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R432), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This work was supported by the Deanship of Scientific Research, Vice President for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant KFU241063].

REFERENCES

[1] L. Aggarwal, P. Goswami, and S. Sachdeva, "Multi-criterion intelligent decision support system for COVID-19," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107056.

[2] R. Blazek, L. Hrosova, and J. Collier, "Internet of Medical Things-based clinical decision support systems, smart healthcare wearable devices, and machine learning algorithms in COVID-19 prevention, screening, detection, diagnosis, and treatment," *Amer. J. Med. Res.*, vol. 9, no. 1, pp. 65–80, 2022.

[3] N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi, and J. S. Suri, "Human activity recognition in artificial intelligence framework: A narrative review," *Artif. Intell. Rev.*, vol. 55, no. 6, pp. 4755–4808, Aug. 2022.

[4] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa, "Deep learning in human activity recognition with wearable sensors: A review on advances," *Sensors*, vol. 22, no. 4, p. 1476, Feb. 2022.

[5] M. M. Rahman, K. Beng Gan, and N. A. Aziz, "A review on challenges in telerehabilitation and human activity recognition approaches during COVID-19 pandemic," *Jurnal Kejuruteraan*, vol. 35, no. 3, pp. 577–586, May 2023.

[6] Q. Li, T. You, J. Chen, Y. Zhang, and C. Du, "LI-EMRSQL: Linking information enhanced Text2SQL parsing on complex electronic medical records," *IEEE Trans. Rel.*, vol. 73, no. 2, pp. 1280–1290, Jun. 2024, doi: 10.1109/TR.2023.3336330.

[7] N. Wang, J. Chen, W. Chen, Z. Shi, H. Yang, P. Liu, X. Wei, X. Dong, C. Wang, L. Mao, and X. Li, "The effectiveness of case management for cancer patients: An umbrella review," *BMC Health Services Res.*, vol. 22, no. 1, Oct. 2022, doi: 10.1186/s12913-022-08610-1.

[8] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. D. Munari, "IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8553–8562, May 2019.

[9] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, p. 2556, Nov. 2017.

[10] K. D. Garcia, T. Carvalho, J. Mendes-Moreira, J. M. Cardoso, and A. C. de Carvalho, "A study on hyperparameter configuration for human activity recognition," in *Proc. Int. Workshop Soft Comput. Models Ind. Environ. Appl.* Cham, Switzerland: Springer, 2019, pp. 47–56.

[11] K. Dorgham, H. Ben-Romdhane, I. Nouaouri, and S. Krichen, "A decision support system for smart health care," in *IoT and ICT for Healthcare Applications* (EAI/Springer Innovations in Communication and Computing), N. Gupta and S. Paiva, Eds. Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-030-42934-8_6.

[12] X. Zhou, W. Liang, K. I. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-learning-enhanced human activity recognition for Internet of healthcare things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6429–6438, Jul. 2020.

[13] Z. Qin, Y. Zhang, S. Meng, Z. Qin, and K.-K.-R. Choo, "Imaging and fusing time series for wearable sensor-based human activity recognition," *Inf. Fusion*, vol. 53, pp. 80–87, Jan. 2020.

[14] H. Mliki, F. Bouhlel, and M. Hammami, "Human activity recognition from UAV-captured video sequences," *Pattern Recognit.*, vol. 100, Apr. 2020, Art. no. 107140.

[15] Q. Jiang, X. Fu, S. Yan, R. Li, W. Du, Z. Cao, F. Qian, and R. Grima, "Neural network aided approximation and parameter inference of non-Markovian models of gene expression," *Nature Commun.*, vol. 12, no. 1, p. 2618, May 2021.

[16] N. Rashid, B. U. Demirel, and M. A. Al Faruque, "AHAR: Adaptive CNN for energy-efficient human activity recognition in low-power edge devices," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13041–13051, Aug. 2022.

[17] I. U. Khan, S. Afzal, and J. W. Lee, "Human activity recognition via hybrid deep learning based model," *Sensors*, vol. 22, no. 1, p. 323, Jan. 2022.

[18] A.-A. Nahid, N. Sikder, and I. Rafi, *KU-HAR: An Open Dataset for Human Activity Recognition*. London, U.K.: Mendeley, Feb. 2021, doi: 10.17632/45F952Y38R.5.

[19] E. Bisong, "Introduction to Scikit-learn," in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Berkeley, CA, USA: Apress, 2019, doi: 10.1007/978-1-4842-4470-8_18.

[20] D. Khan, M. Alonazi, M. Abdelhaq, N. Al Mudawi, A. Algarni, A. Jalal, and H. Liu, "Robust human locomotion and localization activity recognition over multisensory," *Frontiers Physiol.*, vol. 15, Feb. 2024, Art. no. 1344887, doi: 10.3389/fphys.2024.1344887.

[21] Y. Lin, C. Chen, Z. Ma, N. Sabor, Y. Wei, T. Zhang, M. Sawan, G. Wang, and J. Zhao, "Emulation of brain metabolic activities based on a dynamically controllable optical phantom," *Cyborg Bionic Syst.*, vol. 4, p. 47, Jan. 2023, doi: 10.34133/cbsystems.0047.
[22] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.
[23] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Forecasting of COVID-19 using deep layer recurrent neural networks (RNNs) with gated recurrent units (GRUs) and long short-term memory (LSTM) cells," *Chaos, Solitons Fractals*, vol. 146, May 2021, Art. no. 110861.
[24] S. Mekruksavanich and A. Jitpattanakul, "CNN-based deep learning network for human activity recognition during physical exercise from accelerometer and photoplethysmographic sensors," in *Proc. ICCBI* Singapore: Springer, 2021, pp. 531–542.
[25] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020.
[26] Y. Zhao, S. Chen, S. Liu, Z. Hu, and J. Xia, "Hierarchical equalization loss for long-tailed instance segmentation," *IEEE Trans. Multimedia*, vol. 26, pp. 6943–6955, 2024, doi: 10.1109/tmm.2024.3358080.
[27] W. Zheng, S. Lu, Y. Yang, Z. Yin, and L. Yin, "Lightweight transformer image feature extraction network," *PeerJ Comput. Sci.*, vol. 10, p. e1755, Jan. 2024, doi: 10.7717/peerj-cs.1755.
[28] C. Yang, D. Sheng, B. Yang, W. Zheng, and C. Liu, "A dual-domain diffusion model for sparse-view CT reconstruction," *IEEE Signal Process. Lett.*, vol. 31, pp. 1279–1283, 2024, doi: 10.1109/lsp.2024.3392690.
[29] S. R. S. Chakravarthy and H. Rajaguru, "Detection and classification of microcalcification from digital mammograms with firefly algorithm, extreme learning machine and non-linear regression models: A comparison," *Int. J. Imag. Syst. Technol.*, vol. 30, no. 1, pp. 126–146, Mar. 2020.
[30] S. R. S. Chakravarthy, N. Bharanidharan, and H. Rajaguru, "Multi-deep CNN based experimentations for early diagnosis of breast cancer," *IETE J. Res.*, vol. 69, no. 10, pp. 7326–7341, Oct. 2023.
[31] M. H. Abid and A.-A. Nahid, "Two unorthodox aspects in handcrafted-feature extraction for human activity recognition datasets," in *Proc. Int. Conf. Electron., Commun. Inf. Technol. (ICECIT)*, Sep. 2021, pp. 1–4.
[32] M. H. Abid, A.-A. Nahid, M. R. Islam, and M. A. P. Mahmud, "Human activity recognition based on wavelet-based features along with feature prioritization," in *Proc. IEEE 6th Int. Conf. Comput., Commun. Autom. (ICCCA)*, Dec. 2021, pp. 933–939.
[33] P. Kumar and S. Suresh, "DeepTransHHAR: Inter-subjects heterogeneous activity recognition approach in the non-identical environment using wearable sensors," *Nat. Acad. Sci. Lett.*, vol. 45, no. 4, pp. 317–323, Aug. 2022.
[34] S. Mekruksavanich, P. Jantawong, N. Hnoohom, and A. Jitpattanakul, "ResNet-based network for recognizing daily and transitional activities based on smartphone sensors," in *Proc. 3rd Int. Conf. Big Data Anal. Practices (IBDAP)*, Sep. 2022, pp. 27–30.
[35] I. D. Luptáková, M. Kubovčík, and J. Pospíchal, "Wearable sensor-based human activity recognition with transformer model," *Sensors*, vol. 22, no. 5, p. 1911, Mar. 2022.

S. R. SANNASI CHAKRAVARTHY received the B.E. degree in electronics and communication engineering (ECE) from P. T. Lee CNCET, Kanchipuram, India, in 2010, the M.E. degree in applied electronics from the Thanthai Periyar Institute of Technology, Vellore, India, in 2012, and the Ph.D. degree in machine learning from Anna University, Chennai, in 2020. He is currently an Associate Professor with the Bannari Amman Institute of Technology, India. His research interests include machine learning and deep learning.

N. BHARANIDHARAN received the Ph.D. degree from the Faculty of Information and Communication Engineering, Anna University, Chennai, in 2020. He is currently an Assistant Professor with the School of Computer Science Engineering and Information Systems (SCORE), Vellore Institute of Technology, Vellore, India. His research interests include machine learning and deep learning techniques for solving various real-world problems.

V. VINOTH KUMAR (Member, IEEE) is currently an Associate Professor with the School of Computer Science Engineering and Information Systems (SCORE), Vellore Institute of Technology (VIT), Vellore, India. He is the author/coauthor of papers in international journals and conferences, including SCI-indexed papers. He has published more than 60 papers in IEEE Access, Springer, Elsevier, IGI Global, and Emerald. His current research interests include wireless networks, the Internet of Things, machine learning, and big data applications. He is an Associate Editor of *International Journal of e-Collaboration* and *International Journal of Pervasive Computing and Communications* and an editorial member of various journals.

T. R. MAHESH (Senior Member, IEEE) is currently an Associate Professor and the Program Head with the Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bengaluru, India. He has to his credit more than 100 research articles in Scopus and SCIE-indexed journals of high repute. He has been an editor for books on emerging and new-age technologies with publishers, such as Springer, IGI Global, and Wiley. He has served as a reviewer and a technical committee member for multiple conferences and journals of high reputation. His research interests include image processing, machine learning, deep learning, artificial intelligence, the IoT, and data science.

SURBHI BHATIA KHAN (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering, specialized in machine learning and social media analytics. She is currently with the Department of Data Science, School of Science, Engineering and Environment, University of Salford, Manchester, U.K. She earned Project Management Professional Certification from the reputed Project Management Institute, USA. She has more than 13 years of academic and teaching experience at different universities. She has published many papers in reputable journals and conferences in high-indexing outlets. Her research interests include machine learning, sentiment analysis, and data science.

AHLAM ALMUSHARRAF received the Ph.D. degree in business administration, specializing in information systems. She is currently an Assistant Professor with the College of Business and Administration, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include AI, IS applications, social media, e-commerce, and ICT.

EID ALBALAWI is currently an Assistant Professor with King Faisal University, Saudi Arabia. He has published many journal articles, book chapters, and conference papers in various internationally recognized academic databases. He is contributing to the research community by various volunteer activities. His research interests include evolutionary computation, reinforcement learning, and AI.