Hybrid CNN-XGB Framework for Enhancing Human Activity Recognition

Farah Hatem Khorsheed*[1], Raniah Hazim[1], Sarah. A. hassan[1], Qusay Saihood[2]
[1]Department of computer engineering, Collage of engineering, University of Diyala, Iraq
[2]Prosthetic Dental Techniques Department, College of Health and Medical Techniques, Ashur University, Baghdad, Iraq
Emails: farah_hatam@uodiyala.edu.iq; Rania_hazim_enge@uodiyala.edu.iq; Sarah.amir@uodiyala.edu.iq; qusaysaihood@au.edu.iq

Abstract

Human Activity Recognition (HAR) is one of the most important modern research fields concerned with studying and analyzing human actions and behaviors. Human activity recognition applications offer great potential for a wide range of applications in various fields that enhance health, safety, and efficiency. Due to the diversity of human activities and the way people carry out these activities, it is difficult to recognize human activity. The amazing capabilities provided By Artificial Intelligence (AI) tools in analyzing and understanding hidden patterns in complex data can greatly facilitate the HAR process. There has been a huge trend in the past 10 years to use Machine Learning (ML) and Deep Learning (DL) techniques to analyze and understand big data for HAR. Although there are many studies using these techniques, their accuracy still needs to be further improved due to several challenges: Data complexity, class imbalance, determining the appropriate feature selection technique with ML technique, and tuning the hyperparameters of the used ML technique. To overcome these challenges, this study proposes an effective framework based on two stages: a data preprocessing procedure that includes data balance and data normalization. Then, a hybrid CNN-XGB model combining Convolutional Neural Network (CNN) and a fine-tuned XGBoost (XGB) classifier is developed for accurate HAR. The CNN-XGB model achieved excellent results in HAR when trained and tested on the HCI-HAR dataset, achieving an accuracy of up to 99.0%. Effectively HAR provides the opportunity to apply many applications that contribute to improving the quality of life in various areas of our daily lives.

Keywords: Human Activity Recognition; Machine Learning; Deep Learning; Convolutional Neural Network; XGBoost; HCI-HAR dataset.

1. Introduction

HAR has become a key technology with applications in many different industries, such as activity monitoring, assisted living, smart home systems and fall detection [1]. The Global institute on aging huge budget annually ($4,412.1 million for 2024 budget) supports research and resources to facilitate behavioral and social research on the life course, aging, and Alzheimer's disease and Alzheimer's disease-related dementias (AD/ADRD) globally. HAR provides ample space for the elderly by maintaining their independence, health, and activity levels, thereby reducing the economic impacts of global aging. The number of research articles on HAR has increased over the past ten years, indicating the growing interest in this field and its many applications in fields including virtual reality, security cameras, elder care, and human-computer interaction. The growing recognition of HAR's potential impact on a variety of sectors is shown in this increased trend in academic articles [2]. The field's dynamic nature is driven by developments in sub-topics, changes in funding, and the continuous introduction of innovative methods.

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

Tracking daily activities provides vital information in various aspects that can be used to create applications that keep pace with rapid technological development [3]. The daily activities of humans are broad and cannot be confined within limits due to the great diversity of movements and activities that humans practice in their daily lives. For example, the activities inside a shopping centre could be fall under selling, buying, eating, playing with children, arranging goods, changing clothes, playing sports, etc. Therefore, events can be predicted according to the places where people are present. Regardless of the endless activities and behaviour in various situations, there are basic activities such as walking, sitting, lying down, climbing the stairs, descending the stairs, and standing that people practice in any place and time. Tracking and monitoring these activities provide insight into individuals' daily physical activity levels and overall health. Monitoring a person going up and down the stairs can provide information about the health of the heart and blood vessels and give early indicators that could save the individual’s life from a medical standpoint. As well as, monitoring the movement of prolonged lying or sitting can give a reading about weight gain and obesity in terms of physical fitness and exercise. These activities are also necessary to monitor and evaluate the progress made in the rehabilitation of patients recovering from injuries and surgeries. The most important aspect from our point of view is detecting falls in the elderly, especially in the case of elderly individuals who live alone [4].

Whenever it is mentioned above, it underscores the imperative to delve into this aspect comprehensively, leveraging all material and moral resources to foster its development and capitalize on its potential benefits. However, neglecting it and failing to pay attention means we overlook the chance to harness its practical and technical advantages. It is impossible to track human daily activities visually or aurally, so the process of monitoring began via sensors that are installed through chips and implanted in the human body or by wearing wearable measuring devices [2] [3]. In the same regard, from a practical standpoint, it is also not possible to collect, store and analyse the data and information collected through these sensors manually, so it requires the introduction of AI techniques to solve this problem, these technologies have proven highly efficient in analysis, classification, and prediction, as well as in making decisions very quickly.

The use of computer vision to identify human activity has broad applications, especially in the medical field where it helps streamline elderly patients' emergency response and motion rehabilitation monitoring [5]. Applications in education, athletics, and even social relationships with robots are possible. Novel products like smart cameras, wearable technology, health and fitness applications, automated surveillance systems, and human-computer interface systems have been made possible by recent developments in HAR [6].

While several ML and DL techniques have been developed to tackle the challenges of human action recognition, each one still has its own shortcomings, advantages, and disadvantages. The analysis, depiction, and detection of human actions are the main goals in HAR [4]. Although classic ML techniques have been widely applied to accomplish these goals, they frequently lack the stability necessary for precise action recognition. Technological progress in the last several years has made it easier to train and use CNN frameworks that can learn representations directly from data, removing the requirement for hand-crafted features/rules. This is especially important when more data becomes available, and these frameworks get more accurate very quickly [7].

Despite the application of various ML and DL techniques in the HAR sector, there are still some challenges facing the application of these techniques due to the dynamic nature of this sector, which can be summarized as follows:

- Given the nature of the data, which is in the form of time series, it is difficult to identify key features using traditional feature selection methods.
- There is no specific technique for human activity recognition that is significantly superior to other techniques.
- The accuracy of ML and DL techniques used for HAR still needs improvement.
- Existing datasets often suffer from an imbalance between classes, and this creates a problem for any ML technique.

The main contributions of this study are:

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

- To Solve the problem of unbalanced classes within the data set by using the Synthetic Minority Over-sampling Technique (SMOTE) to increase the elements of the minority classes.
- To development of a hybrid CNN-XGB model that combines a CNN for feature extraction and selection and a fine-tuned XGB classifier for an accurate HAR.
- The performance of the proposed model was evaluated by calculating accuracy, precision, recall, and F1 score.
- The results obtained from this study were compared with the state-of-the-art. In the same regard, A Receiver Operating Characteristic (ROC) curve showing the true positive rate versus false positive rate was plotted to classify each category of human activity, in addition to calculating the area under the ROC curve (AUC) for each category of human activities.

Finally, the rest of this paper is divided as follows: Section 2 summarizes related work, Section 3 gives an overview of the AI techniques used. Section 3 explains the workflow of the proposed framework. The results are presented and discussed in Part 4. In the last section, conclusions, and future work.

2. Related Work

In the past few years, there has been a great trend in using various ML and DL techniques in HAR Applications. Activity recognition utilizing Hidden Markov Models (HMMs) to capture temporal relationships between activities is the focus of this paper, which was written by Licheng Zhang, Xihong Wu, and Dingsheng Luo. The emission distribution of HMMs was typically represented by generative models, such as Gaussian mixture models (GMMs), or discriminative models, such as Random Forest (RF). Nevertheless, all techniques need manual feature extraction, which is a laborious process that might lead to data loss during sensor quantization. The authors suggest a unique method for modeling the emission distribution of HMMs that makes use of deep neural networks (DNNs). DNNs do not require human feature extraction; instead, they automatically extract features appropriate for classification from raw sensor data. The HMM-DNN model was tested against HMM-GMM and HMM-RF using a dataset of everyday activities. The outcomes show that HMM-DNN outperformed HMM-GMM and HMM-RF in performance, while the results were 0.93, 0.84 and 0.86 [8].

An architecture for HAR based on smartphone inertial accelerometer is presented by Wan et al. This architecture solves the shortcomings of conventional techniques in recognizing complex and real-time human actions with multimodal and high-dimensional sensor data in the context of mobile edge computing (MEC). The suggested approach entails gathering sensory data sequences from subjects performing everyday activities, obtaining high-efficiency features, and applying a CNN-based real-time human activity classification technique. Using the UCI and Pamap2 datasets, the study evaluates the performance of several models. The outcomes show how effective the suggested strategy is on these huge public datasets. The UCI dataset achieved the best results with the CNN 93.2 % [9].

Seungmin Oh, Akm Ashiquzzaman, Dongsu Lee, Yeonggwang Kim, and Jinsul Kim discuss their research on HAR utilizing Semi-Supervised Active Transfer Learning. The study addresses the issues of constructing DL models for HAR that require a large amount of labeled data. Manual labeling in domains such as HAR is costly and time-consuming, resulting in sluggish data collection and sometimes biased labeling. This study presents an in-depth analysis of HAR that investigates recent developments in computer vision using CNN. The suggested technique reduces labeling efforts on new data by exploiting information learned from semi-supervised active transfer learning. This strategy outperformed random sampling and typical active transfer learning methods by 95.9%, with minimal labeling efforts [10].

Serpush and Rezaei [11] tackled the preprocessing challenge by automatically selecting illustrative frames from the input stream. Instead of using the whole feature set, they take the method of extracting significant properties from the frames. The suggested hierarchical method makes use of Histogram of Oriented Gradients (HOG) and Background Subtraction (BGS), which are combined with DNN (Deep Neural Networks) and skeletal modeling. In addition, the method uses CNN and Long Short-Term Memory (LSTM) recursive networks for feature selection to maintain earlier data. Finally, Softmax-KNN classifier helps in the classification of human actions. This study used The UCF101 dataset includes 101 action classes, over 13,000 video clips and 27 hours of video data, we compare with the five actions that we need for our study. The named the method Hierarchical Feature Reduction & DL"-based action recognition method which achieved 83.2% accuracy.

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

Tasnim et al. [12] developed a spatial-temporal image formation (STIF) technique for three-dimensional (3D) skeleton joints that uses both spatial data and temporal alterations to discriminate actions. The authors extracted discriminative features from ImageNet datasets using Transfer Learning (TL) and pretrained algorithms such as DenseNet121, ResNet18, and MobileNetV2. Furthermore, they evaluated the provided methodology using several fusion methods. The study investigated the effects of three fusion models—maximization, elementwise average, and multiplication—on the efficiency fluctuations of HAR. They used UTD-MHAD (University of Texas at Dallas multi-modal human action dataset) and MSR-Action3D (Microsoft action 3D) as a dataset where the accuracy of testing was 98.9 %.

Jaouedi et al. [7] proposed a unique HAR approach based on feature extraction and video analysis. The method integrates motion information derived from human action tracing using Gaussian Mixture Models (GMM) and Kalman Filter (KF) methods. Furthermore, features based on each visual characteristic across all video sequence frames are extracted using the Recurrent Neural Network (RNN) approach, which employs a Gated Recurrent Unit (GRU). The main benefit of this novel technology is its capacity to extract and evaluate each characteristic continuously across all instances and frames in the video. The proposed approach is evaluated on the challenging UCF Sports, UCF101 and KTH datasets. An average of 96.3% is obtained when we have tested on KTH dataset.

Yu et al. [13] used ensemble DL to analyze body postures and recognize backdrop data in images to accomplish HAR. Using the pretrained CNN method, the authors developed an end-to-end CNN technique. The capacity of CNN to learn independent spatial and channel information using parallel branches improves the technique's overall efficiency. As a result, they developed an end-to-end DELWO technique that takes advantage of a nonsequential topology to manually combine deep data from several sources. Finally, the authors developed the DELVS model, which combines numerous deep techniques and weighted coefficients to produce optimal forecasts. They conducted experiments in Li's action dataset, uncropped and 1.5x cropped Willow action datasets while the results were 96.6 by ResNet50_CNN model, 66.0 by VGG19_CNN model and 62.8 by ResNet50_CNN model, respectively.

Sargano et al. [14] proposed a unique HAR approach that uses pretrained CNN models as a fundamental framework to extract features from a target dataset. A hybrid Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) technique is then used to categorize actions. The authors demonstrated the effective transportability of previously learned CNN-based representations obtained from large-scale annotated datasets to HAR challenges with limited training datasets. UCF sports and KTH datasets were used while the results by SVM-KNN model achieved 98.6% in accuracy.

Mathe et al. [15] proposed a CNN-based technique for HAR that focuses on Activities of Daily Living (ADL). The approach uses Discrete Fourier Transform (DFT) images to train the CNN model. The authors created 3D skeleton positions for human joints from raw Red, Green, and Blue (RGB) signals and improved them with depth data. In addition, 3-1D signals were collected to characterize the mobility of each joint, giving its coefficient in 3D Euclidean regions. The developers then tested the proposed model on a difficult, publicly available dataset that includes one or more parts of the human body to classify daily activities.

Mukherjee et al, [16] presented a model based on convolutional pooling, which employed a simple convolutional network with LSTM, Encoding-net, and CNN-net to create three classifiers that differed in kernel size, dense layers, and main variance in the framework. The main idea of each classifier is for converting time series data into a two-dimensional matrix by opening a window to predict types of human behavior. After that, the three classifiers are combined through the general framework for the purpose of polling on the results based on the product rule method, majority voting, merging the results, as well as the adaptive weighted method, known as the group rule, the study used three different types of data, and the highest results were obtained by applying the UCF11 database, with a rate of 98.2%.

Xia et al. [17] used a hybrid model that included LSTM and CNN for activity recognition. The LSTM module collected temporal information from sequential multimodal mobile sensor data, while the CNN module extracted essential characteristics. The suggested model attained an F1 score of 95.78% by optimizing hyperparameters, including batch normalization, and using the UCI-HAR dataset, outperforming other baseline models.

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

Yan Sun, et al [18]. had created a LaptSNE which is a novel graph-layout nonlinear dimensionality reduction approach based on t-SNE that is one of the best strategies for visualising high-dimensional data as 2D scatter plots. The approach is assessed through a formal comparison with state-of-the-art algorithms on seven benchmark datasets (Waveform, PenDigits, COIL 20, COIL 100, HAR, MNIST, and Fashion-MNIST). The findings suggest that the proposed technique is superior to baselines such as t-SNE and UMAP.

Moharana, B. et al. [19] believed that monitoring human activity is a crucial undertaking with applications in marketing, healthcare, surveillance, and other fields. They therefore trained an artificial neural network and a long short-term memory neural network to categorize the activities into different groups. They then compared these neural networks to each other and to the previously used methods & algorithms on the dataset. The findings of their evaluation of the protected model using the Human Activity Recognition Dataset from the UCI ML Repository, which is made up of information gathered from the cellphones of thirty participants, were quite encouraging.

Li, Xue, et al. [20] they proposed using inverse CNN techniques to improve the extraction and selection of key features. The performance of the inverse CNN is compared with both traditional CNN and CNN-LSTM on five datasets commonly used in HAR, including the HCI-HAR dataset. The accuracy of the inverse CNN was 97.12% when applied to the HCI-HAR dataset.

3. Overview of Used Techniques

3.1. Convolutional Neural Networks (CNNs)

CNNs are a subclass of DNNs which represent the essential component to computer vision applications since they are specifically made for processing and evaluating visual data. CNNs utilize specialized layers to automatically extract hierarchical characteristics from visual data, thereby revolutionizing computer vision [9]. Pooling layers down sample spatial dimensions, fully connected layers combine high-level features for decision-making, and convolutional layers apply filters to discover local patterns Figure 1 [21]. Non-linearity is introduced by activation functions, and supervised learning is used in backpropagation to optimize the model [16]. CNNs are crucial to AI because of their distinctive architecture annon-linearityich have made them useful for tasks like image identification.

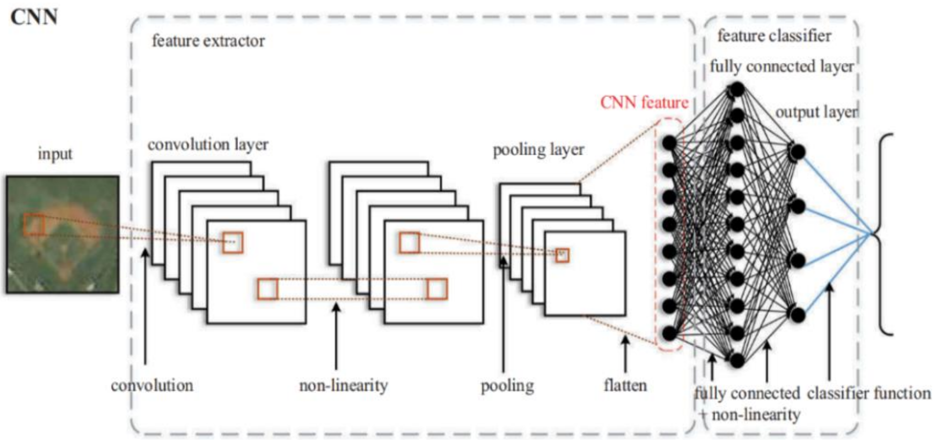

Figure 1: CNN network Architecture.

3.2. XGBoost (XGB) Classifier

Extreme Gradient Boosting, or XGBoost, is a potent ensemble learning algorithm renowned for its accuracy and efficiency. It is based on a gradient boosting architecture and use regularization to avoid overfitting. Shallow decision trees are used as base learners as shown in Figure 2 [22].XGB is a popular choice for classification and regression tasks in both small and large-scale applications since

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

it evaluates feature importance, facilitates cross-validation, and allows for parallel and distributed computing. XGB based on the theory of classification and regression tree can be a highly effective method of regression and classification problems [19],[17].

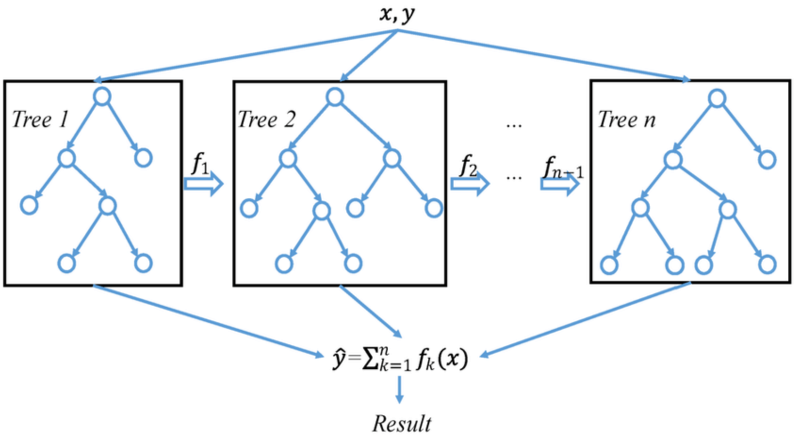

Figure 2: General Architecture of XGB Classifier.

4. The proposed Framework

This work proposes a hybrid framework for human activity recognition based on two stages as shown in Figure 3: In the first stage, data pre-processing is performed which includes class balancing within HCI-HAR dataset and data normalization. In the second stage, a hybrid model is developed that combines CNN network with a well-tuned XGB ensemble classifier for accurate recognition of human activities.

4.1. Dataset

The HCI-HAR is a Human Activity Recognition database created by recording 30 study participants conducting daily activities (ADL) while wearing a waist-mounted smartphone with inbuilt inertial sensors. The goal is to classify actions into one of six categories. A group of thirty participants between 19 and 48 participated in the trials. Every individual carried out six tasks (Walking, Walking Downstairs, Walking Upstairs, Sitting, Standing, Laying) while sporting a Samsung Galaxy S II smartphone around their waist. A three-dimensional linear acceleration and three-dimensional angular velocity have been recorded at a steady rate of 50 Hz using its integrated accelerometer and gyroscope. To manually identify the data, the experiments were videotaped. After the dataset was acquired, it was randomly divided into two sets, with 70% of the volunteers chosen to create training data and 30% to create test data [23].

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

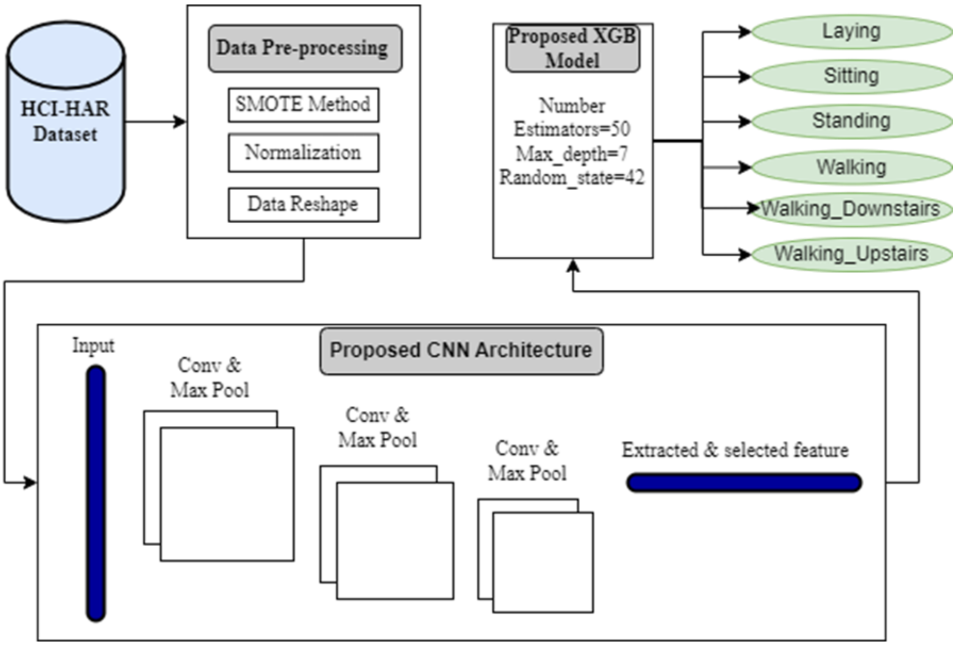

Figure 3: The general framework of the proposed hybrid CNN-XGB model for human activity recognition.

4.2. Data Pre-processing

Data pre-processing in ML and DL is important because the performance of these models depends on the quality of the data fed to it. At this stage, the data used is usually cleaned, transformed, and organized into a format suitable for the ML techniques used [24]. In this work, the following pre-processing steps are performed on the HCI-HAR dataset: data integration, class balancing, data normalization, and data transformation.

The HCI-HAR dataset consists of two files, one file containing the training set and the other file containing the test set, and they have been combined into one file to make the complete number of data 10299. In the second stage, the Synthetic Minority Over-sampling Technique (SMOTE) was used, it is a technique to increase the data of minority groups within a data set to balance the distribution of data [25]. Table 1 shows the distribution of data by the six categories of human activities within the HCI-HAR data set after and before using SMOTE.

In the third stage, the data was normalized using Min-Max normalization to standardize the data range to have a common range between 0 and 1, according to the equation below:

$$X\ Normlized\ = \frac{Original\ X - Min\ X}{Max\ X - Min\ X} \qquad (1)$$

Where the X normalized refers to the normalized value, original X refers to the original value, Min X refers to the minimum value, Max X refers to the maximum value.

In the final stage, the values of the category column are converted into a numeric representation using one-Hot encoding technique to fit CNN network.

Table 1: The Distribution of HCI-HAR data set after and before using SMOTE.

The Class Name	Number of samples before using SMOTE	Number of samples after using SMOTE
Laying	1944	1944
Sitting	1777	1944

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

Standing	1906	1944
Walking	1722	1944
Walking_Downstairs	1406	1944
Walking_Upstairs	1544	1944

4.4. Hybrid CNN-XGB Model

After completing the data pre-processing stage, the data is entered into the proposed hybrid CNN-XGB Model based on the fusion between the CNN network and the XGB classifier. The CNN network works to reduce the dimensions of the used HCI-HAR dataset, which determines the most key features affecting the classification process, so that the classification process can be carried out effectively using the well-tuned XGB classifier.

The CNN architecture developed in this study consists of three one-dimensional convolutional layers (Conv1D) and three one-dimensional maximum pooling layers (MaxPooling1D), with the ReLU activation function used for all convolutional layers. In addition to a Flatten layer to transform the data output from the previous convolutional and pooling layers into a vector form as shown in Table 2.

Table 2: Shows the layers and the outputs of proposed CNN network.

Layer Name	Output Shape	Param
Conv1D	(560, 128)	512
MaxPooling1D	(280, 128)	0
Conv1D	(278, 64)	24640
MaxPooling1D	(139, 64)	0
Conv1D	(137, 32)	6176
MaxPooling1D	(68, 32)	0
Flatten	(2176)	0

Finally, the outputs of the CNN network are entered as input to the well-tuned XGB classifier to classify human activities with high efficiency. The performance of the XGB model was tuned by adjusting the hyperparameters that are important for this model. The number of trees used for the XGB classifier was set to 50, in addition to the depth of each tree set to 7.

4.5. Evaluation

The multiple confusion matrixes are defined as a tool used to evaluate the performance of an intelligent system if it is classified in more than one category and measure its quality. The general shape of the matrix is a square with two dimensions that correspond to the total number of target categories. As for this matrix's cells, the expected categories appear within them for both its columns and rows. In addition to the use of the confusion matrix, the ROC curve and the AUC curve were used, and they are also among the most important measurement tools used. In the end, all the (Accuracy, Precision, Recall, and F1 Score) were displayed in the evaluation and results section.

5. Results and Discussion

The proposed model was developed using the TensorFlow and xgboost libraries, in addition to the Scikit Learn library for data pre-processing and model evaluation. The proposed model was trained and tested on the HCI-HAR dataset, where the dataset was separated using the train-test-split method into 80% training set and 20% test set.

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

Performance of experiments indicates the proposed CNN-XGB Model achieved promising results in recognizing human activities with a high accuracy of up to 99.0%.

The model's performance was evaluated by recognizing each activity, where the proposed CNN-XGB model achieved precision, recall, and F1 score of 100.0% in recognizing the activity of lying down and Walking Downstairs. Also, as shown in Table 3 CNN-XGB model achieved precision, recall, and F1 score of up to 99.0% in recognizing the activity of Walking and Walking Upstairs. while the results of recognizing sitting and standing activity were up to 97.0% for both precision, recall, and F1 score.

Table 3: The results of Proposed Model for each human activity in HCI-HAR dataset.

Category	Precision	Recall	F1-Score
Laying	100.0%	100.0%	100.0%
Sitting	97.0%	97.0%	97.0%
Standing	97.0%	97.0%	97.0%
Walking	99.0%	99.0%	99.0%
Walking_Downstairs	100.0%	100.0%	100.0%
Walking_Upstairs	99.0%	99.0%	99.0%
Macro Average	99.0%	99.0%	99.0%
Weighted Average	99.0%	99.0%	99.0%

A multiple confusion matrix is also calculated to evaluate the proposed model, as shown in Figure 4. The proposed model achieved completely correct recognition for all Laying activity samples, while there was only one incorrect recognition for the Walking Downstairs activity out of a total of 382 samples when tested on the unseen data of the HCI-HAR dataset, as shown in Figure 4. While out of 385 samples for the Sitting category, there were 10 samples that were incorrectly recognized by the proposed model, and out of 386 samples for the Standing category, there were 11 samples that were also incorrectly recognized.

A ROC curve showing the true positive rate versus false positive rate has been plotted to recognize each category of human activities as shown in Figure 5. In addition to calculating the AUC for each category of human activities, where AUC for the proposed CNN-XGB model was 100% for all categories except for two categories (Sitting and Standing), which reached 98%.

Finally, the above-mentioned results when evaluating the proposed model using various performance metrics indicate that the model's performance was 100% accurate when recognizing Laying and Walking Downstairs activities. The model also performed very well in recognizing Walking and Walking Upstairs activities, while the weakest performance of the model was observed in recognizing Sitting and Standing activities.

This study Proved that balancing classes within a dataset, as well as combining appropriate ML techniques and tuning the hyperparameters of these techniques, can help in HAR with high efficiency. The hybrid CNN-XGB model proposed in this study showed significant superiority over the ML techniques used in the studies discussed in the related work section, as shown in Table 4.

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

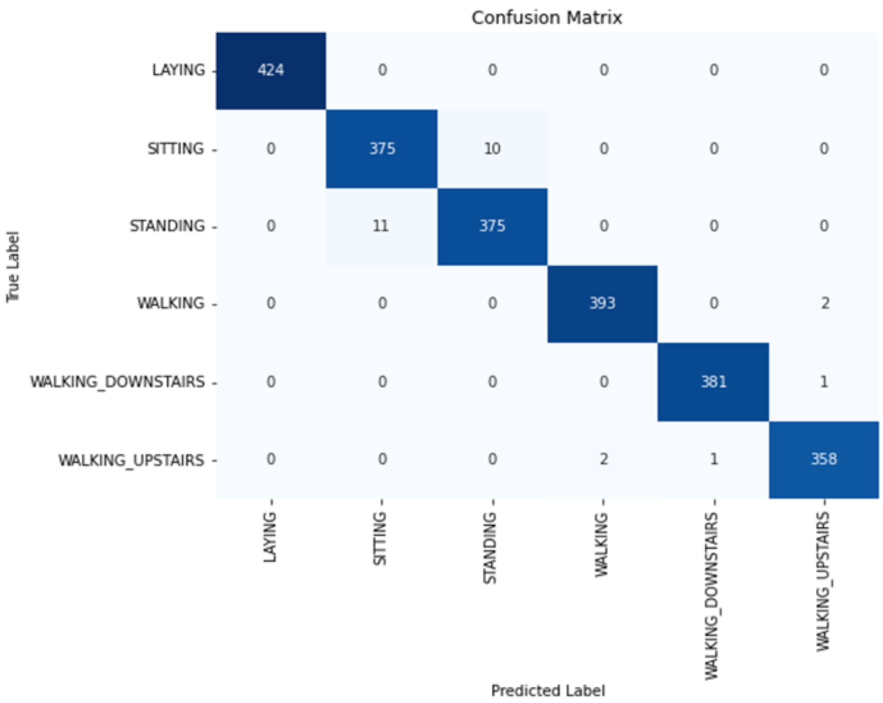

Figure 4: The Multiclass Confusion Matrix of Proposed CNN-XGB Model.

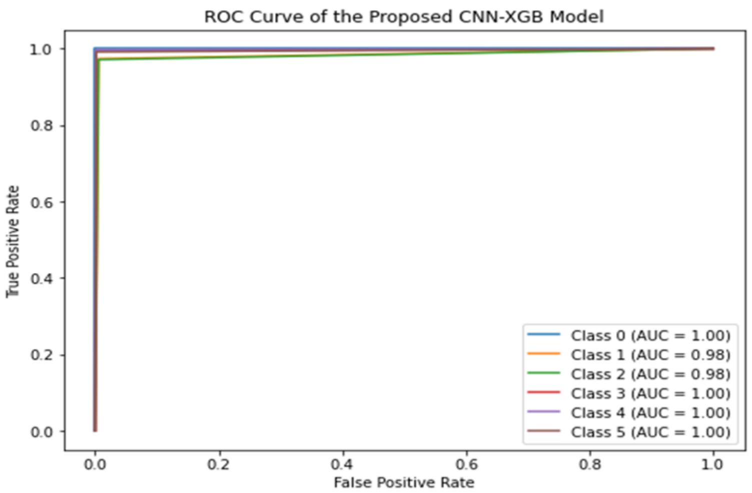

Figure 5: The ROC curve of Proposed CNN-XGB Model for Each human activity in HCI-HAR dataset

Table 4: Comparing the proposed model with previous work.

The Study	Accuracy
LaptSNE [18]	93.0%
ANN Architecture [19]	95.05%
Reversed CNN [20]	97.12%

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

Our Hybrid CNN-XGB Model	99.0%

6. Conclusion

This study proposed a hybrid framework based on CNN and XGB classifier called CNN-XGB Model to improve human activity recognition. First, data pre-processing was performed, which included data balancing, data normalization, and data transformation. Then the hybrid CNN-XGB model was used for a more effective recognition process. Performance experiments were conducted on the HCI-HAR dataset, which contains 6 different human activities, and the proposed CNN-XGB model achieved promising results in recognizing these activities. The proposed CNN-XGB model gives an accuracy of up to 99.0% in the recognition process, superior to the traditional models of ML and DL that were mentioned in previous studies. For future work, the authors plan to test the proposed model on different datasets to identify human activities. There is also a fantastic opportunity to use pre-trained CNN models instead of traditional CNN models to improve the results and reduce the time in tuning the performance of the CNN model.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Saeed, V. A. (2024). A Framework for Recognition of Facial Expression Using HOG Features. International Journal of Mathematics, Statistics, and Computer Science, 2, 1–8. https://doi.org/10.59543/ijmscs.v2i.7815.

[2] H. M. Salman, H. F. Hamdan, R. Khalid, and S. Al Al-Kikani, “Physical Activity Monitoring for Older Adults through IoT and Wearable Devices: Leveraging Data Fusion Techniques,” Fusion: Practice and Applications, vol. 11, no. 2, pp. 48–61, 2023, doi: 10.54216/FPA.110204.

[3] .. R., .. S. S., and A. Sharma, “Predicting Student Performance Using Educational Data Mining and Learning Analytics Technique,” Journal of Intelligent Systems and Internet of Things, vol. 10, no. 2, pp. 24–37, 2023, doi: 10.54216/JISIoT.100203.

[4] Zeebaree, I. M., & Kareem, O. S. (2024). Face Mask Detection Using Haar Cascades Classifier To Reduce The Risk Of Coved-19. International Journal of Mathematics, Statistics, and Computer Science, 2, 19–27. https://doi.org/10.59543/ijmscs.v2i.7845.

[5] Y. Zhang, Z. Deng, X. Xu, Y. Feng, and S. Junliang, “Application of Artificial Intelligence in Drug–Drug Interactions Prediction: A Review,” J Chem Inf Model, vol. 64, no. 7, pp. 2158–2173, Apr. 2024, doi: 10.1021/acs.jcim.3c00582.

[6] K. Shukla, K. Idanwekhai, M. Naradikian, S. Ting, S. P. Schoenberger, and E. Brunk, “Machine Learning of Three-Dimensional Protein Structures to Predict the Functional Impacts of Genome Variation,” J Chem Inf Model, Apr. 2024, doi: 10.1021/acs.jcim.3c01967.

[7] N. Jaouedi, N. Boujnah, and M. S. Bouhlel, “A new hybrid deep learning model for human action recognition,” Journal of King Saud University - Computer and Information Sciences, vol. 32, no. 4, pp. 447–453, May 2020, doi: 10.1016/j.jksuci.2019.09.004.

[8] L. Zhang, X. Wu, and D. Luo, “Human activity recognition with HMM-DNN model,” in 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), IEEE, Jul. 2015, pp. 192–197. doi: 10.1109/ICCI-CC.2015.7259385.

[9] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, “Deep Learning Models for Real-time Human Activity Recognition with Smartphones,” Mobile Networks and Applications, vol. 25, no. 2, pp. 743–755, Apr. 2020, doi: 10.1007/s11036-019-01445-x.

[10] M. G. Morshed, T. Sultana, A. Alam, and Y.-K. Lee, “Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities,” Sensors, vol. 23, no. 4, p. 2182, Feb. 2023, doi: 10.3390/s23042182.

[11] F. Serpush and M. Rezaei, “Complex Human Action Recognition Using a Hierarchical Feature Reduction and Deep Learning-Based Method,” SN Comput Sci, vol. 2, no. 2, p. 94, Apr. 2021, doi: 10.1007/s42979-021-00484-0.

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024

[12] N. Tasnim, M. K. Islam, and J.-H. Baek, "Deep Learning Based Human Activity Recognition Using Spatio-Temporal Image Formation of Skeleton Joints," Applied Sciences, vol. 11, no. 6, p. 2675, Mar. 2021, doi: 10.3390/app11062675.
[13] X. Yu et al., "Deep Ensemble Learning for Human Action Recognition in Still Images," Complexity, vol. 2020, pp. 1–23, Jan. 2020, doi: 10.1155/2020/9428612.
[14] Md. Ferdouse Ahmed Foysal, M. Shakirul Islam, S. Abujar, and S. Akhter Hossain, "A Novel Approach for Tomato Diseases Classification Based on Deep Convolutional Neural Networks," 2020, pp. 583–591. doi: 10.1007/978-981-13-7564-4_49.
[15] E. Mathe, A. Maniatis, E. Spyrou, and P. Mylonas, "A Deep Learning Approach for Human Action Recognition Using Skeletal Information," 2020, pp. 105–114. doi: 10.1007/978-3-030-32622-7_9.
[16] Z. Zhang, Z. Lv, C. Gan, and Q. Zhu, "Human action recognition using convolutional LSTM and fully-connected LSTM with different attentions," Neurocomputing, vol. 410, pp. 304–316, Oct. 2020, doi: 10.1016/j.neucom.2020.06.032.
[17] S. Satpathy, P. Mohan, S. Das, and S. Debbarma, "A new healthcare diagnosis system using an IoT-based fuzzy classifier with FPGA," J Supercomput, vol. 76, no. 8, pp. 5849–5861, Aug. 2020, doi: 10.1007/s11227-019-03013-2.
[18] Y. Sun, Y. Han, and J. Fan, "Laplacian-based Cluster-Contractive t-SNE for High-Dimensional Data Visualization," ACM Trans Knowl Discov Data, vol. 18, no. 1, pp. 1–22, Jan. 2024, doi: 10.1145/3612932.
[19] B. Moharana et al., "Human Activity Recognition Using Artificial Neural Networks and Long Short-Term Memory Networks," *Sensors*, vol. 22, no. 15, p. 5200, 2022.
[20] X. Li, Y. Zhang, and Z. Wang, "Improved Feature Selection for Human Activity Recognition Using Inverse CNN," *Journal of Machine Learning*, vol. 15, no. 4, pp. 505-518, 2021.
[21] Vikas panthi, Amit Kumar Mishra, Enhancing Healthcare Monitoring through the Integration of IoT Networks and Machine Learning, Journal of International Journal of Wireless and Ad Hoc Communication, Vol. 7 , No. 1 , (2023) : 28-39 (Doi : https://doi.org/10.54216/IJWAC.070103).
[22] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," Geoscience Frontiers, vol. 12, no. 1, pp. 469–477, Jan. 2021, doi: 10.1016/j.gsf.2020.03.007.
[23] "ESANN 2013 - proceedings | ESANN 2024." Accessed: Apr. 19, 2024. [Online]. Available: https://www.esann.org/proceedings/2013
[24] Q. Saihood and E. Sonuç, "The Efficiency of Classification Techniques in Predicting Anemia Among Children: A Comparative Study," 2022, pp. 167–181. doi: 10.1007/978-3-030-97255-4_12.
[25] Q. SAIHOOD and E. SONUÇ, "A practical framework for early detection of diabetes using ensemble machine learning models," Turkish Journal of Electrical Engineering and Computer Sciences, vol. 31, no. 4, pp. 722–738, Jul. 2023, doi: 10.55730/1300-0632.4013.

Doi: https://doi.org/10.54216/FPA.150218
Received: August 02, 2023 Revised: December 09, 2023 Accepted: April 07, 2024