Mediterranean Smart Cities Conference (MSCC 2024)
2-4 May, 2024, Martil, Morocco

Enhancing LSTM Models for Human Activity Recognition in Health Smart Homes with SMOTE and Hyperparameter Tuning

Mariam El Ghazi
Information Technology and Modeling, Systems Research Unit, Abdelmalek Essaadi University, Tetuan, Morocco
mariam.elghazi@etu.uae.ac.ma

Noura Aknin
Information Technology and Modeling, Systems Research Unit, Abdelmalek Essaadi University, Tetuan, Morocco
noura.aknin@uae.ac.ma

Abstract— **Health Smart Homes (HSH) have become essential for enhancing the autonomy and privacy of elderly and dependent individuals, thanks to recent advancements in IoT sensor technologies. These homes require reliable Human Activity Recognition (HAR) devices for precise monitoring. However, collecting sensor data may result in imbalanced data across different activities due to variations in their durations. This imbalance presents a challenge for Deep Learning (DL) models, as they struggle to handle such disparate data effectively, impacting their ability to recognize and generalize patterns within diverse activity classes accurately.**

This research explores the influence of class imbalance on multiple DL models in HAR using the PAMAP2 Dataset. Initially, a baseline experiment is conducted by training the Long Short-Term Memory (LSTM) model on the imbalanced Dataset. Subsequently, a second experiment is carried out on balanced data using the Synthetic Minority Over-sampling Technique (SMOTE). For both experiments, Hyperparameter Tuning was performed to achieve the optimal performance of the model. The findings underscore the importance of Dataset balance, leading to notable enhancements in critical metrics such as accuracy and F1 score. The SMOTE sampling technique is a practical approach to address this issue. This study emphasizes the significance of tackling class imbalance in Human Activity Recognition using DL models and presents a practical solution. The successful implementation of the SMOTE technique holds promising implications for enhancing the accuracy and performance of HSH Systems, furthering elderly care, increasing their independence, and protecting their privacy. Future research will delve into alternative Datasets, advanced models, and diverse sampling techniques to continually refine HAR capabilities.

Keywords— Deep Learning (DL), Health Smart Homes (HSH), Human Activity Recognition(HAR), SMOTE, Class Imbalance, Hyperparameter Tuning.

I. INTRODUCTION

The anticipated rise in the global population aged 60 years and above, reaching 2 billion by 2050, according to UNFPA [1], poses a significant challenge for healthcare systems. This demographic shift necessitates continuous healthcare monitoring, placing immense stress on medical caregivers and traditional hospital-based methods. The conventional healthcare approach is costly and inefficient in handling the growing demand [2]. As a viable alternative, implementing health monitoring systems in smart home environments emerges as a solution. This approach offers real-time healthcare support for individuals seeking to maintain their independence, alleviating pressure on healthcare institutions [3].

The system focuses on four main types of activities: Activities of Daily Living (ADL), encompassing routine tasks like eating and washing, Instrumental Activities of Daily Living (IADL), including community living tasks such as meal preparation and medication use, Ambulatory activities, which are linked to the subject's motion and posture, Lastly physiological activities such as cardiac and brain activities, addressing the comprehensive healthcare needs of the aging population [3].

In the realm of HSH Systems, The emergence of the Internet of Things (IoT) has expanded the sensor landscape, including body-worn devices for monitoring physiological signals, object-attached sensors for capturing movement, ambient sensors integrated into the environment, and hybrid combinations of various sensor types [4].

DL models, notably LSTM and Convolutional Neural Networks (CNN), have significantly transformed HAR by enhancing the effective processing and analysis of sensor data. However, real-world scenarios present challenges in collecting sensor data, mainly due to variations in activity durations, leading to class imbalances [5][6]. This imbalance can adversely impact the effectiveness of HAR models, introducing bias and compromising accuracy [7]. A pivotal strategy to address this challenge involves integrating sampling techniques, such as SMOTE. This method proves critical in augmenting the overall performance of HAR models. This study emphasizes the indispensable role of balanced Datasets in the HAR domain, offering practical insights directly applicable to real-world scenarios.

The primary contributions of this study are :

- Investigate the implications of class imbalance on HAR models in HSH.
- We explore the role of sampling techniques, such as SMOTE, to address the class imbalance problem.
- Perform comparative experiments and results analysis of the impact of SMOTE and Hyperparameter Tuning on the performance of the LSTM-based model.
- Evaluate the model in terms of accuracy and F1 score to assess model performance, offering insights compared to previous studies and contributing to the ongoing refinement of HSH Systems.

The paper is structured into clear sections to present the research findings systematically. Section 2 delves into Related Works, reviewing previous research about the identified problem in the field. Section 3 provides an in-depth exploration of the Materials and Methods utilized in the experimental approach. Moving forward, Section 4 outlines the results of the experiments and conducts a comprehensive discussion of the findings. Lastly, Section 5, the Conclusion, succinctly summarizes the essential findings and outlines potential future directions for research.

II. RELATED WORKS

Several notable studies have significantly advanced HAR using DL models on the PAMAP2 Dataset. Wan et al. (2020) achieved high accuracies of 91.00% and 85.86% with CNN and LSTM models on smartphones, respectively [8]. Xu et al. (2022) introduced classical CNN, LSTM, and Inception-LSTM methods, achieving F1 scores ranging from 0.8949 to 0.9513 [9]. Tehrani et al. (2023) employed a deep Bi-LSTM model, obtaining promising results of 93.41% in both metrics [10]. Thakur et al. developed a CNN and LSTM hybrid model with an autoencoder, reaching an impressive F1 score of 0.9446 and an accuracy of 94.33% [11]. Challa et al. (2022) proposed a CNN-BiLSTM model, achieving an accuracy of 94.29% [12]. Alharbi et al. (2022) explored oversampling methods like SMOTE, showing improvement on the PAMAP2 Dataset, with an MLP achieving an F1 score of 0.7185 compared to the baseline of 0.7473 [13].

Table 1 summarizes the performance of these studies. Our study builds upon these insights, focusing on HAR with the PAMAP2 Dataset.

TABLE 1. PREVIOUS RELATED RESEARCH PERFORMANCE METRICS ON PAMAP2

Study	year	Classification method	Accuracy	F1 score
[8]	2020	LSTM	0.8580	0.8534
[9]	2022	LSTM	0.8920	0.8949
[10]	2023	Bi-LSTM	0.9341	0.9341
[12]	2022	CNN-BiLSTM	0.9429	-
[11]	2022	ConvLSTM AE	0.9433	0.9446
[13]	2022	MLP+ SMOTE	--	0.7473

Despite advancements in HAR, there is a notable gap in the literature concerning applying sampling techniques to address class imbalance. While numerous studies have leveraged DL models for improved HAR accuracy, limited attention has been given to systematically exploring the effectiveness of sampling techniques like SMOTE. This gap highlights the need for a more comprehensive understanding of how such techniques can be strategically integrated to optimize model performance, especially in real-world scenarios where class imbalances may pose challenges. In this study, we will address this gap and explore the impact of sampling techniques like SMOTE on the performance of DL models.

III. MATERIALS AND METHODS

In this study, experiments are conducted to investigate the influence of class imbalance on the performance of DL models for HAR. To tackle this issue, the SMOTE sampling technique is examined, involving training an LSTM model on the PAMAP2 Dataset. The anticipated findings aim to provide valuable insights for improving the accuracy and reliability of HAR systems within HSH.

A. PAMAP2 Dataset

The PAMAP2 dataset is a data collection approach collected from numerous wearable sensors (the **wrist, the chest, and the ankle)** on diverse subjects [14]. Table 2 describes the different features of the PAMAP2 Dataset.

TABLE 2. PAMAP2 DATASET DESCRIPTION

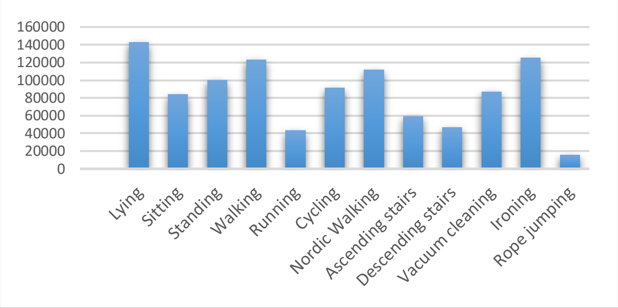
Activity number	Sampling Rate	Windows Size	Ove r- lap	Number of participants
12	100 Hz	1s	50%	9

We observe that the PAMAP2 Dataset exhibits class imbalance attributed to variations in the duration of different activities, as depicted in Table 3 and Figure 1. This imbalance presents a challenge for training accurate DL models.

TABLE 3. INSTANCES NUMBER PER CLASS ACTIVITY IN THE PAMAP2 DATASET

Class Id	Activity Label	Instances	Class Id	Activity Label	Instances
0	Lying	142931	6	Nordic Walking	111832
1	Sitting	83738	7	Ascending stairs	59314
2	Standing	99973	8	Descending stairs	46830
3	Walking	122906	9	Vacuum cleaning	86959
4	Running	43050	10	Ironing	125228
5	Cycling	91340	11	Rope jumping	15453

Figure 1. Number of Instances by Activity in the PAMAP2 Dataset



B. *Synthetic Minority Over-sampling Technique (SMOTE):*

This study uses SMOTE to deal with the problem of class imbalance. This technique proves to be a valuable tool for addressing data imbalance in Datasets. It produces synthetic data for the minority class, linking gaps between existing samples. This process is particularly beneficial for DL models, as it creates a substantial amount of data, ultimately balancing the Dataset and significantly enhancing classification accuracy [15].

C. *Long Short-Term Memory (LSTM)*

One of the most used DL models for HAR is LSTM. This model is a type of RNN (Recurrent Neural Network). LSTMs are used for time series scenarios, particularly HAR. LSTM is mainly used for sensor-based activity classification, such as accelerometers and smartphone gyroscope readings. The unique architecture of LSTMs allows them to capture and process sequential information effectively, making them well-suited for tasks involving temporal dependencies, like those encountered in time series applications [16][17].

D. *Model Configuration*

The choice for evaluating the Dataset with the LSTM-based model is as follows: Research findings suggest that LSTM networks demonstrate enhanced recognition efficiency with their iterative extraction of temporal features [18]. The LSTM-based model architecture comprises one LSTM layer, two Dropout layers, two dense layers, and an output layer for activity classification.

E. *Evaluation Metrics*

Numerous evaluation metrics were used to assess the HAR model performance in the experimentation phase. These metrics included the **accuracy, the F1 score, and the confusion matrix.** These metrics will determine the DL model performance on the Dataset. Table 4 demonstrates the definition of all these metrics [19].

TABLE 4. EVALUATION METRICS

Metric	Formula
Accuracy	$\frac{tp + tn}{tp + tn + fp + fn}$
F1 score	$\frac{2(recall * precision)}{recall + precision}$
Confusion matrix	It is a Two-dimension table of class labels, one represents the current class, and the other represents the predicted one.

IV. EXPERIMENTS AND RESULTS

A. *Experiments:*

This paper's objective is to investigate the impact of SMOTE on the performance of DL models for HAR by addressing the following research questions:

1. *Does the issue of class imbalance affect the performance of the* DL *models in sensor-based Datasets?*
2. *What is the effect of SMOTE on addressing the class imbalance in sensor-based HAR?*
3. *Does using SMOTE optimize the performance of the DL model for HAR?*

The experiments are performed by Google Collab Pro+, using the NVIDIA GPU V100. In this study, two experiments were conducted:

- ***Experiment 1*:** We train and test the deep model on an imbalanced Dataset (Baseline).
- ***Experiment 2*:** We train and test the deep models on the balanced Dataset using the SMOTE sampling technique.

To evaluate deep models on the PAMAP2 Dataset, the experiments involve the following stages such as:

- *Data Collection*: This stage represents collecting raw sensor data from the wearable sensors.
- *Data Preprocessing:* this stage includes data cleaning, noise reduction, and data normalization.
- *Data Class Balancing:* the SMOTE techniques were applied to the training set for balancing the data.
- *Data Segmentation:* Data were segmented into overlapping windows.
- *Model Evaluation:* the model's performance is evaluated on the PAMPA2 Dataset using the performance metrics.

B. *Results:*

In Experiment 1, the deep model was trained on an imbalanced Dataset. In experiment 2, the deep model was trained on a balanced Dataset with SMOTE. Then, the model is evaluated using the performance metrics. Then, a comparative analysis of the previous studies on state of the art will be conducted. The deep model underwent Hyperparameter Tuning using the Keras Tuner, as illustrated in Table 5.

TABLE 5. THE SUMMARIZED HYPERPARAMETERS RESULTS OF THE PERFORMANCE OF THE PROPOSED MODEL ON IMBALANCED AND BALANCED DATA

Hyperparameter	Experiment 1 (imbalanced data)	Experiment 2 (balanced data)
Dense Units	32	64
Lstm Units	32	96
Dropout Rate	0.1	0.1
Optimizer	RMSPROP	RMSPROP
Learning Rate	0.001	0.01
Batch Size	32	32
Epochs	82	77

Table 6 shows the results of the two experiments while comparing the performance of the deep model in terms of accuracy and the F1 score.

TABLE 6. RESULTS OF PERFORMANCE OF DEEP MODEL ON IMBALANCED AND BALANCED DATA

	Experiment 1 Results	*Experiment 2 Results*
Accuracy	0.9257	**0.9499**
F1 Score	0.9250	**0.9503**

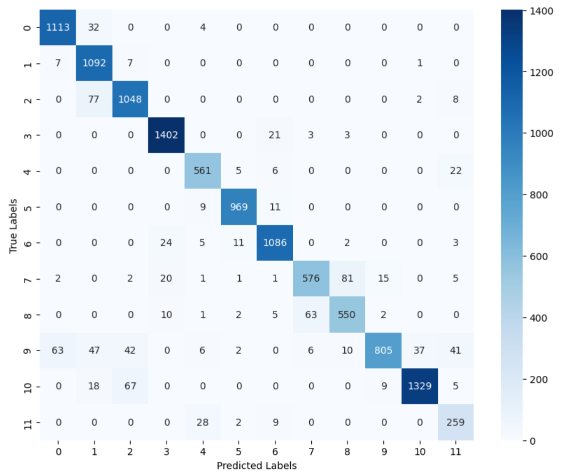


Figure 2. Confusion Matrix of the proposed model on Imbalanced data

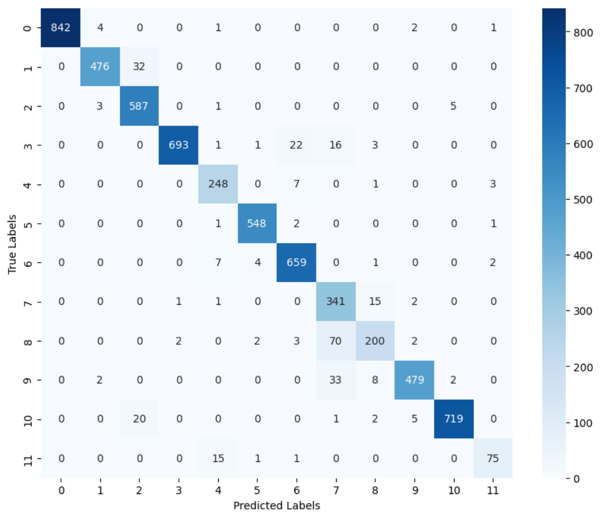


Figure 3. Confusion Matrix Of The Proposed Model On Balanced Data With SMOTE

Figures 2 and 3 show the confusion matrices for the proposed model on the imbalanced and balanced data with SMOTE, respectively.

The state-of-the-art research used several DL models for HAR using the PAMPA2 dataset. Table 7 demonstrates their results; these prior studies have produced impressive outcomes. Finally, this paper demonstrates the effectiveness of SMOTE techniques in addressing the class imbalance in HAR, leading to higher accuracy and F1 scores compared to previous studies, emphasizing the importance of addressing the imbalanced data problem.

TABLE 7. COMPARATIVE RESULTS WITH RELATED WORKS

Study	**Classification method**	**Accuracy**	**F1 score**
[8]	LSTM	0.8580	0.8534
[9]	LSTM	0.8920	0.8949
[10]	Bi-LSTM	0.9341	0.9341
[12]	CNN-BiLSTM	0.9429	-
[11]	convLSTM AE	0.9433	0.9446
[13]	MLP+ SMOTE	--	0.7473
This study	LSTM model on imbalanced data	0.9257	0.9250
This study	**LSTM model And SMOTE**	**0.9499**	**0.9503**

C. Discussion

Previous studies have emphasized the scarcity of research addressing the class data imbalance issue in human activity recognition (HAR) [20], [21]. This study highlights this gap in the research by examining the impact of sampling approaches, specifically SMOTE, on both class imbalance and the performance of HAR models.

Hyperparameter Tuning proved valuable throughout the experiments, ensuring that DL models were optimized for specific conditions. The investigation highlighted SMOTE techniques' impact on DL model performance in HAR. The application of SMOTE, coupled with hyperparameter optimization, resulted in considerable improvements in performance metrics. This highlights the effectiveness of SMOTE, especially when combined with optimal hyperparameters, in addressing class imbalance and enhancing recognition efficiency.

The significant implications of this research extend to healthcare and contribute to enhanced elderly experiences in HSH. Future works could explore additional datasets such as WISDM or Opportunity datasets and more types of DL model architectures and further examine advanced hyperparameter optimization techniques.

V. CONCLUSION

This paper on sensor-based HAR using DL models focuses on enhancing HAR system accuracy, which is crucial for healthcare and sports analytics. Addressing imbalanced Datasets, we employed the SMOTE technique to balance the data and test it with an LSTM model. The results demonstrated significant accuracy and F1 score improvements, particularly in countering class imbalance. As IoT sensors increase, our findings contribute to more intelligent healthcare systems. Future work will involve diverse Datasets, complex models, and additional sampling techniques to advance innovative healthcare systems.

REFERENCES

[1] "Ageing in the Twenty-First Century | Fonds des Nations Unies pour la population." https://www.unfpa.org/fr/node/6225 (accessed Oct. 29, 2021).

[2] F. Sabry, T. Eltaras, W. Labda, K. Alzoubi, and Q. Malluhi, "Machine Learning for Healthcare Wearable Devices: The Big Picture," *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/4653923.

[3] H. Mshali, T. Lemlouma, M. Moloney, and D. Magoni, "A survey on health monitoring systems for health smart homes," *Int. J. Ind. Ergon.*, vol. 66, pp. 26–56, 2018, doi: 10.1016/j.ergon.2018.02.002.

[4] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Feb. 2019, doi: 10.1016/j.patrec.2018.02.010.

[5] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, Jan. 2014, doi: 10.1145/2499621.

[6] F. Alharbi, L. Ouarbya, and J. A. Ward, "Synthetic Sensor Data for Human Activity Recognition," *Proc. Int. Jt. Conf. Neural Networks*, no. July, 2020, doi: 10.1109/IJCNN48605.2020.9206624.

[7] H. F. Nweke, Y. W. Teh, G. Mujtaba, U. R. Alo, and M. A. Al-garadi, "Multi-sensor fusion based on multiple classifier systems for human activity identification," *Human-centric Comput. Inf.*, vol. 9, p. 1, 2019.

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

[9] Y. Xu and L. Zhao, "Inception-LSTM Human Motion Recognition with Channel Attention Mechanism," *Comput. Math. Methods Med.*, vol. 2022, 2022, doi: 10.1155/2022/9173504.

[10] A. Tehrani, M. Yadollahzadeh-Tabari, A. Zehtab-Salmasi, and R. Enayatifar, "Wearable Sensor-Based Human Activity Recognition System Employing Bi-LSTM Algorithm," *Comput. J.*, no. April, 2023, doi: 10.1093/comjnl/bxad035.

[11] D. Thakur, S. Biswas, E. S. L. L. Ho, and S. Chattopadhyay, "ConvAE-LSTM: Convolutional Autoencoder Long Short-Term Memory Network for Smartphone-Based Human Activity Recognition," *IEEE Access*, vol. 10, pp. 4137–4156, Jun. 2022, doi: 10.1109/ACCESS.2022.3140373.

[12] S. K. Challa, A. Kumar, and V. B. Semwal, "A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data," *Vis. Comput.*, vol. 38, no. 12, pp. 4095–4109, 2022, doi: 10.1007/s00371-021-02283-3.

[13] F. Alharbi, L. Ouarbya, and J. A. Ward, "Comparing Sampling Strategies for Tackling Imbalanced Data in Human Activity Recognition," *Sensors*, vol. 22, no. 4, pp. 1–20, 2022, doi: 10.3390/s22041373.

[14] A. Reiss and D. Stricker, "Creating and benchmarking a new Dataset for physical activity monitoring," in *ACM Int. Conf. Proceeding Ser*, 2012.

[15] A. A. A. A. Alani, G. Cosma, and A. Taherkhani, "Classifying Imbalanced Multi-modal Sensor Data for Human Activity Recognition in a Smart Home using Deep Learning," in *Proceedings of the International Joint Conference on Neural Networks*, 2020. doi: 10.1109/IJCNN48605.2020.9207697.

[16] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/NECO.1997.9.8.1735.

[17] M. El Ghazi and N. Aknin, "Optimizing Deep LSTM Model through Hyperparameter Tuning for Sensor-Based Human Activity Recognition in Smart Home," *Inform.*, vol. 47, no. 10, pp. 109–122, 2023, doi: 10.31449/inf.v47i10.5268.

[18] S. Mekruksavanich and A. Jitpattanakul, "Lstm networks using smartphone data for sensor-based human activity recognition in smart homes," *Sensors*, vol. 21, no. 5, pp. 1–25, 2021, doi: 10.3390/s21051636.

[19] N. S. Khan and M. S. Ghani, *A Survey of Deep Learning Based Models for Human Activity Recognition*, vol. 120, no. 2. Springer US, 2021. doi: 10.1007/s11277-021-08525-w.

[20] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Comput. Surv.*, vol. 54, no. 4, Jul. 2021, doi: 10.1145/3447744.

[21] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Comput. Surv.*, vol. 52, no. 4, 2019, doi: 10.1145/3343440.