Modelling Activities of Daily Living Using Local Interpretable Model-Agnostic Explanation Algorithm

Henry Onyekwe, Idongesit Ekerete *, Matias Garcia-Constantino, Chris Nugent.

School of Computing, Ulster University Belfast, BT15 1ED, UK

* Correspondence: i.ekerete@ulster.ac.uk

Abstract—**The use of Artificial Intelligence (AI) in healthcare, particularly in recognising anomalous behaviour during Activities of Daily Living (ADLs), is useful for supporting independent living. Transparency and interpretability of ADLs can play a vital role in decision-making processes, particularly in healthcare sectors. This work intends to offer additional information to AI-based prediction of ADLs through the use of Local Interpretable Model-agnostic Explanations (LIME). In this study, 5,125 low resolution thermal images gleaned from ADLs in a laboratory environment which mimics a smart home were clustered and analysed using Data Mining software and AI algorithms respectively. Results indicated that LIME presented saliency maps of ADLs in diverse scenarios such as 'Making Tea' and 'Sitting Down' to consume it. Further work will seek to fine-tune the models for better accuracy.**

Keywords— Activities of daily living, LIME, Explainable AI, Healthcare, Thermal sensing

I. INTRODUCTION

In recent years, the use of Artificial Intelligence (AI) algorithms in modelling healthcare datasets has presented avenues for activity detection, particularly in the area of detecting abnormal behaviour during the performance of Activities of Daily Living (ADLs) [1]. This development has also presented opportunities for ensuring the safety and well-being of individuals, in independent and in elderly care settings [2]. Even so, as AI systems become more advanced, there is a pressing need for transparency and interpretability of the decision-making processes, particularly in fields where these decisions directly affect the health and quality of life of individuals [1]. This will further enhance the trustworthiness and transparency of the AI model implemented.

Activity Detection and Prediction (ADP) have been broadly studied in many spheres ranging from Engineering, and Computer Science to Health Sciences [3]–[5]. ADP can involve the use of diverse sensing solutions and algorithms. Whilst the former can be implemented with the use of wearable, intrusive and non-intrusive sensing solutions, the latter can employ edge and wavelet scattering algorithms, ensemble learning, Machine Learning (ML) and Deep Learning (DL) algorithms [6]–[10].

AI which embodies ML and DL has offered many sophisticated models in recent years [11]–[13]. Although health-related and complicated activities can be predicted with these models, the transparency and the explainability of their predictions are often questioned [14], [15]. This happens because most of the AI models utilise Blackbox algorithms which put users in doubt about how predictions culminate [1], [16].

The present work aims to enhance the clarity of AI prediction and aid the interpretability of abnormal behaviour detection through the systematic application of Explainable AI (XAI) such as Local Interpretable Model-agnostic Explanations (LIME). Furthermore, this work aims to implement a robust and dependable system that would not only identify abnormal behaviours but also provide clear and understandable justifications for its predictions. The present work also intends to enthrone trust and confidence among stakeholders such as carers, medical professionals, and family members, who rely on monitoring technologies and sensing solutions in healthcare and hospital settings.

The remainder of this paper is organised as follows. Section II discusses related work on ADLs monitoring and prediction. Section III discusses the Methodology used in this work, Section IV presents the Results, and Section V presents the Discussion. Conclusions and References are presented in Sections VI and VII, respectively.

II. RELATED WORK

ADLs monitoring and classification are currently applied in many contexts and studies incorporating the use of wearable sensing solutions and unobtrusive ambient sensing solutions [17]–[19]. These studies also demonstrated the use of data mining models and AI algorithms including sensor fusion technologies. The fusion of datasets from these sensing solutions can help to improve healthcare monitoring practices by providing multimodal sensing avenues, thus ensuring the safety and well-being of individuals in home-based environments. However, these studies did not incorporate explainability and interpretability in their models.

The concept of Explainable Artificial Intelligence (XAI) addresses the need for transparency and comprehensibility, which is one of the obstacles to deploying AI models in healthcare settings [20]. This section reviews key publications that contribute towards the understanding of abnormal behaviour detection during ADLs and highlights the central role of XAI in improving the interpretability and reliability of AI models.

The ability to understand the justifications behind AI-driven decisions becomes essential in situations where those decisions have a direct impact on people's safety and well-being [1]. In the field of healthcare, decisions can have life-altering consequences. It becomes evident that the need for XAI is not only a matter of trust but also a crucial factor for ethical and legal compliance. This sheds light on the unique considerations and complexities of implementing XAI in healthcare settings [21].

Techniques applied in the XAI domain include Local

979-8-3503-5298-6/24/$31.00 ©2024 Crown

Interpretable Model-agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), amongst others [1]. LIME is a prominent XAI algorithm that approximates the predictions of a complex model at a local level and provides clear justifications for specific predictions. Contrariwise, SHAP is a well-known XAI method that gives input features equal credit for model's prediction. This enables a better understanding of feature extraction. Decision trees, rule-based systems, and model-agnostic techniques are also effective tools for interpretable modelling, as they represent binary decisions based on features and enable stakeholders to trace the path of predictions. These techniques are not limited to a particular machine learning model and can be applied to a variety of algorithms, making them versatile instruments for enhancing interpretability [20].

For effective XAI, there are unique challenges associated with accurately identifying abnormal behaviours [1]. These include human behaviour variability, context sensitivity, sensor limitations and noise. Others include privacy concerns, adaptation to individual profiles, dynamic environments, handling multimodal datasets, and interpreting anomalies in medical datasets. Age, health condition, and individual habits all contribute to the diversity and variability in human behaviour. Identifying abnormal behaviour requires an understanding of the activity's context, which can be difficult due to the possibility of noise components or datasets errors. Privacy concerns are also important, as balancing the collection of sufficient data for accurate detection. Adapting to individual profiles, dynamic environments, and the management of multimodal datasets are additional obstacles. Understanding these details is crucial for designing a system that not only excels in precision but also stands out for its interpretability and transparency. Our study is poised to contribute in enhancing the safety and well-being of individuals in healthcare settings with this foundational knowledge [22].

Kamal et al. [23] used the study of Alzheimer's patient data to show how XAI techniques could change the way healthcare diagnoses. Their research demonstrates the revolutionary potential of XAI in medical diagnostics. By utilising XAI methodologies, they provided clear justifications for AI predictions, increasing transparency and empowering healthcare professionals. This research goes beyond Alzheimer's diagnostics, laying the groundwork for a broader application of XAI in the healthcare domain. In another study, by Sharma and Kaur [24], it is emphasised the significance of XAI in complex contexts, particularly in the fields of healthcare and monitoring of ADLs. The research highlights the advantages of integrating XAI into monitoring operations and anomaly detection. It offers significant contributions to the understanding of techniques to improve the interpretability of models and demonstrates how they can be implemented in practical situations. The research emphasises the importance of transparent AI methods in such contexts, given the extensive ramifications that model predictions can cause. The work of Sharma and Kaur serves as inspiration for the practical implementation of explainability in complex environments. This promotes transparent decision-making processes and accurate predictions, which in turn foster trust among stakeholders and facilitate widespread adoption [24]. The 2018 survey by Adadi and Berrada offers an exhaustive synopsis of XAI approaches and methodologies [23]. Their work informs about several ways to make it easier to understand how "black boxes" work in AI models. These include model-specific interpretability techniques, post-hoc explanations, and rule-based systems, which are all described in more detail in the survey. The survey results provided valuable information that forms the basis of the project, facilitating an all-encompassing comprehension of the tools that can be utilised to improve the interpretability of AI models [12].

XAI deployment faces issues such as balancing interpretability, model performance, and regulatory compliance. Building trust necessitates honest explanations, yet models that are highly interpretable may sacrifice performance. Balancing explainability with ethical concerns is equally difficult. Despite this, XAI has regulatory compliance possibilities [13]. The objective of the present work includes to: (i) cluster thermal images obtained during ADLs, (ii) utilise distinct AI algorithms such as CNN and Keras to examine the grayscale thermal images, (iii) compare the performance of the models, (iv) utilise LIME to obtain saliency maps of the datasets, and (v) enable anomaly detection systems to provide transparent justifications for their predictions.

III. METHODOLOGY

To implement this multifaceted research, a comprehensive methodology was followed. First, datasets gleaning from a low-resolution thermal camera comprising diverse ADLs were utilised. The rationale for using low resolution grayscale datasets was to preserve the privacy of participants. This dataset, which mimics real life ADLs, serves as the foundation for identifying and classifying activities carried out in home settings.

Furthermore, the datasets were clustered using the Orange data mining software to help identify different stages of the ADLs. The stages considered and the number of datasets involved are as shown in Table 1.

Table 1. ADLs stages and datasets involved in the study.

S/No.	ADLs Stages	Datasets
1	Making Tea	866
2	Entering And Leaving Kitchen	248
3	Hot Water Only	991
4	Sitting Down	2159
5	Walking Around Kitchen	519
6	Boiling Water	342
	Total	5125

The datasets in Table 1 presented a case of imbalanced data with the stage 'Sitting Down' having more than a third of the datasets (2,159), and the stage 'Entering And Leaving Kitchen' having the least as 248. This is a common problem in ML that could result in class separation and poor result in model performance amongst others [25]. Hence, one of the limitations of this study.

After the datasets are clustered according to the ADLs stages, a Convolutional Neural Network (CNN) model was crafted as a benchmark and then, a LIME algorithm. The latter presents an advanced toolkit for elucidating specific predictions, thereby providing further explanations to the predicted activities. This integration is anticipated to provide granular insights into decision-making processes in health-related settings. In addition, it provides mapping and transparency to the predictions.

Furthermore, this work employs a mixed-methods approach which includes quantitative and qualitative data collection and analysis such as data pre-processing, model construction, XAI integration, evaluation, and comparative analysis. The rationale for using a mixed-methods approach is justified by its effectiveness in addressing the research problem and harmonising with the study's objectives [26]. This methodology enables a comprehensive investigation of data-driven insights, model creation, and the incorporation of XAI methodologies. It allows for complete evaluation and comparison, ensuring the robustness and interpretability of abnormal behaviour detection in the ADLs context.

Utilising unobtrusive sensing technologies such as thermal sensors, the research team gathered data in a laboratory setting that simulated a smart home environment. The involvement of participants in the ADLs tasks contributed to the acquisition of a real-life dataset. By employing a hierarchical clustering strategy and the Orange data mining software, the data was sorted using classification-by-clustering [4], [27] of the thermal images. The ADLs were carried out with minimal disruption to daily activities, hence, a vivid representation of real-life activities [28].

Data Pre-processing involved a series of stages aimed at optimising the learning process of the machine learning model. The initial stage involved loading the data, which includes extracting image data features and labels from the datasets. The greyscale images were then read using OpenCV command line before the images were scaled to 128x128 pixels. Normalisation was used to keep pixel values consistent to improve the model's learning process. The processed data was stored in the data list and annotated in the labels list. Following that, the data was converted to arrays for efficient computation and manipulation during model training. In summary, the pre-processing steps encompasses standardisation of image dimensions, normalization of pixel values, and the division of the dataset into training and testing subsets, thus laying the groundwork for the development and testing using AI models.

Model development utilised CNN model for detecting abnormal behaviour during ADLs. CNNs were selected because of their capacity to understand hierarchical patterns and perform well in image-based analysis [29]. The model is built with the Keras framework, which includes convolution, pooling, flattening, and fully connected layers. The CNN model was compiled and optimised using the Adam optimizer and sparse categorical cross-entropy loss function. Recursive Auto-Encoders (RAEs) was investigated as an additional tool for detecting anomalies, recognising temporal connections, and sequential patterns in data [30].

Furthermore, unlike CNN model, LIME is used for local interpretability, ensuring transparency in labelling certain behaviours as abnormal. Moreover, LIME helps the interpretability and transparency of abnormal behaviour forecasts. The mixed-methods approach is appropriate for tackling the research challenge and guaranteeing trust in AI systems in healthcare, notably in the setting of abnormal behaviour detection during ADLs.

IV. Results

To evaluate the results fully, we investigated many elements relating to the performance, predictions, and visualisation of the models employed for abnormal behaviour identification during ADLs. These included the performance of CNN, Keras and LIME.

A. Performance Evaluation of CNN

Precision, recall, and F1-score metrics were used to test the accuracy of the CNN model as presented in Table 2.

Table 2: Classification report for CNN

Class Label	Precision	Recall	F1-Score
Making Tea	0.71	0.74	0.72
Entering And Leaving Kitchen	0.50	0.02	0.04
Hot Water Only	0.84	0.99	0.91
Sitting Down	0.96	1.00	0.98
Walking Around Kitchen	0.45	0.37	0.40
Boiling Water	0.30	0.33	0.34
Overall Accuracy	0.80		

From Table 2, the stages of 'Making Tea', 'Hot Water Only' and 'Sitting Down' (see Figures 1 – 3) each achieved an accuracy of more than 70%. The model also achieved an overall accuracy of 80% in identifying anomalous behaviours in a variety of tasks carried out in the kitchen area. However, similar activities such as 'Entering and Leaving Kitchen', 'Walking Around Kitchen' and 'Boiling Water' attained lower accuracy, as presented in Figures 4 – 6.

Figure 1. Making Tea during ADLs

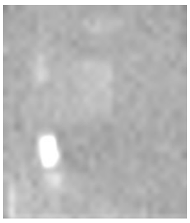


Figure 2. Hot Water Only ready for Tea/Coffee

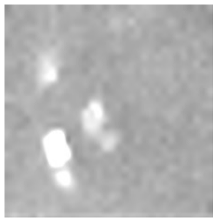


Figure 3. Sitting Down to Consume Tea/Coffee

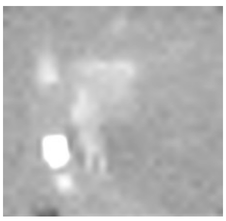


Figure 4. Entering and Leaving Kitchen

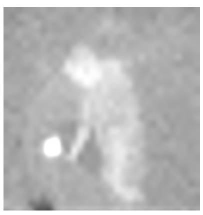


Figure 5. Walking Around Kitchen

Figure 6. Boiling Water during ADLs

B. *Performance Comparison of CNN and Keras (VGG16)*

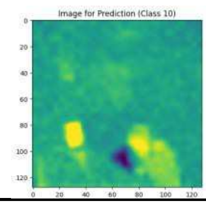
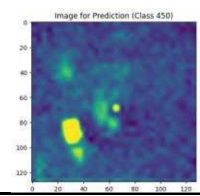
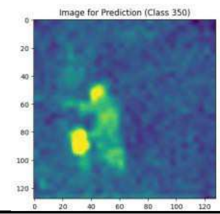
Using the previously mentioned metrics to compare Keras (VGG16) and the custom CNN model performance during ADLs, classification results are presented in Table 3.

Table 3: Classification score for Keras and CNN

Metrics	Keras (VGG16)	CNN
Accuracy	0.4068	0.8000
Precision	0.1655	0.7767
Recall	0.4068	0.8000
F1-Score	0.2353	0.7769

In Table 3, the overall accuracy of CNN doubles the accuracy value of Keras. A similar trend is observed in precision, Recall and F1-Score. Although Keras used VGG16, a deep CNN architecture designed for image classification, its low accuracy metrics may be attributed to a compromise between the reduction of false positives and the augmentation of true positives. Another area of comparison of the performance of CNN with Keras models is the labels appended to the activities as presented in Table 4.

Table 4: Prediction carried out On CNN and Keras

Image	CNN Prediction	Keras Prediction
	Making Tea	Sitting Down
	Sitting Down	Sitting Down
	Boiling Hot Water	Sitting Down

From Table 4, CNN predicted the ADLs such as 'Making Tea', 'Sitting Down' and 'Boiling Hot Water' correctly whilst the VGG16-based fine-tuned Keras model was incapable of predicting the ADLs other than 'Sitting Down' due to its model efficiency. Furthermore, incorrect representation of diverse behaviours within the training data may have also contributed to incorrect classifications. Due to these factors, the model was unable to effectively differentiate rather generalise, resulting in an excessive dependence on specific patterns and the misinterpretation of the ADLs.

B. *Analysis of LIME Visualisations*

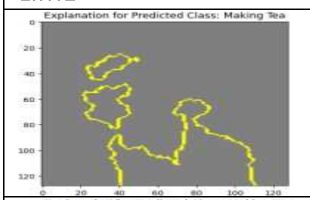
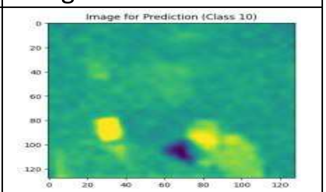
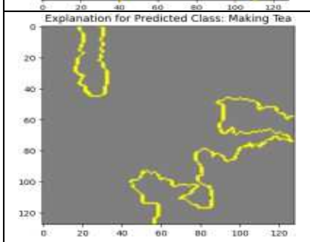
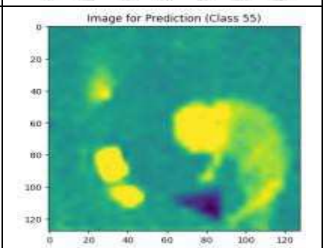
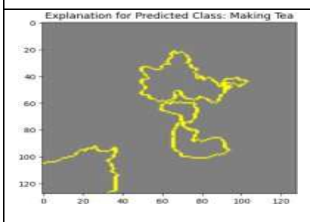
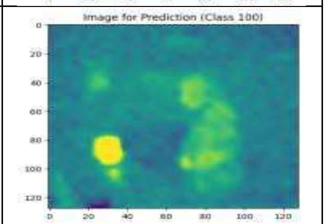
LIME functions by emphasising local interpretability, wherein predictions are explicated at the instance level as opposed to the complete dataset [1]. This approach modifies input features to understand predictions across multiple models [31]. It approximates the original model within a specified neighbourhood by generating simplified local models around particular instances. LIME emphasises the significance of features through the utilisation of weighted samples, saliency maps counterfactuals and feature attribution [32]. The visualisation of LIME's results improves the interpretability of the model by assisting users in comprehending influential features for a specific prediction [33]. This method enhances the credibility of AI model judgements by providing clarity regarding their decision-making procedure [33].

LIME employs relative importance scores rather than standardised units to quantify the significance of features via weighted samples [34]. Computed using the weights assigned to perturbed samples produced in the vicinity of an instance of interest, these scores indicate the importance of features in the context of a particular prediction. LIME functions behave as comparative metrics, prioritising the significance of various attributes in shaping the locally interpretable model developed for a specific case. These scores are, nevertheless, relative indicators of feature prominence due to the absence of a standardised unit of measurement.

C. *CNN LIME Visualisation Comparison*

LIME prediction of ADLs such as 'Making Tea' can produce saliency maps of the thermal blobs in the instance of thermal camera as was the case in this study. C omparisons of LIME visualisations for three predictions in this class are presented in Table 5.

Table 5: LIME Visualisation of ADLs predictions.

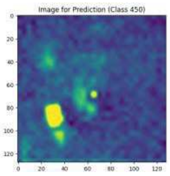
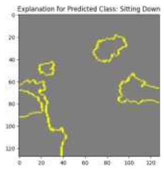
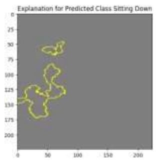
LIME	Images
	
	
	

In Table 5, the visualisation maps depict the individual's ADLs patterns as they evolve with time. In determining the probability of ‘Making Tea’, this visualisation employs comprehensive ADLs patterns as its foundation, rather than singling out any particular isolated action. The maps cover the areas of prediction for the images depending on the thermal blobs spread in the images. All three visualisations seem to affirm the prediction of ‘Making Tea’ by showing the areas of the image that are instrumental to the prediction.

D. Keras and CNN LIME Visualisation Comparison

LIME visualisation using Keras and CNN of a single activity, ‘Sitting Down’ showed further insights as presented in Table 6.

Table 6. LIME visualisation for ‘Sitting Down’.

Prediction	Image	CNN	Keras
Sitting Down			

In Table 6, interesting observations were made after employing LIME to assess the interpretability of the CNN and Keras models. Similar regions were covered by the Keras LIME and CNN LIME explanations, indicating that the two explanations shared a certain degree of consistency. Although the accuracies of these explanations were not computed, the prediction by CNN seems to offer a close semblance to the original image in row 2 column 2 (see Table 6). This analysis emphasises the capacity of CNN to incorporate more complex feature details, while LIME demonstrates its ability to provide coherent insights into model decisions.

V. Discussion

Performance differences between a custom CNN and Keras (VGG16) models were found to be statistically significant when attempting to identify abnormal behaviours in domestic activities. The accuracy of the CNN model was 80%, whereas Keras performed less effectively at 40.68%, predicting exclusively 'Sitting Down' accurately. Additionally, accuracy, recall, and F1- score were better with the CNN model. The similarities between the two models were underscored through the LIME visualisation, which emphasised CNN's capacity to incorporate intricate features. Consistent with the research conducted by Barr Kumarakulasinghe et al. [35] regarding the interpretability of LIME, this study demonstrated its efficacy in elucidating model predictions. In general, the CNN model exhibited a good performance compared to Keras.

Comparing this study to a previous study on evaluating LIME on clinical machine learning classification models, levels of overlap and satisfaction are determined [3]. Their study examines the applicability of LIME as a tool for interpreting black-box machine learning models within healthcare environments. Their results indicated that LIME explanations correspond to the interpretations of clinicians in several instances, demonstrating their medical relevance and hence the importance of their application to ADLs. Nevertheless, concerns related to trust suggest that LIME would not invariably enthrone trust in AI predictions rather explanation which can invoke trust.

VI. Limitation of the Study

Although LIME interpretability of ADLs datasets was relevant in this study, imbalanced data (earlier mentioned) adversely affected its accuracy. Furthermore, ADLs stages such as 'Entering and Leaving Kitchen', ‘Boiling Water’ and 'Walking Around Kitchen,' achieved lower accuracy results due to the similarities of these activities and probably also, due to the size of their datasets.

VII. Conclusion

In summary, the performance disparities between the custom CNN and Keras (VGG16) models in detecting anomalous behaviours in kitchen activities were found to be significant during the comparative analysis. In comparison to Keras, the CNN model demonstrated increased accuracy, precision, recall, and F1-Score metrics. The LIME visualisation drew attention to parallels between the models while emphasising CNN's capacity to incorporate more intricate characteristics. On the other hand, Keras capability to comprehend a wide range of behaviours can be achieved by enhancing the model architecture, augmenting, and diversifying the training datasets, and optimising feature extraction techniques. In addition, further investigation, and application of sophisticated interpretability techniques such as LIME would be beneficial for gaining a comprehensive understanding of model predictions. For real-world applications, research concentrating on enhancing interpretability and model performance in complex behavioural recognition tasks would be invaluable. Future work would also involve finetuning the models to achieve better accuracy.

VIII. References

[1] S. Khedkar, V. Subramanian, G. Shinde, and P. Gandhi, “Explainable AI in Healthcare,” *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3367686.

[2] A. Arcelus, R. Goubran, M. H. Jones, and F. Knoefel, “Integration of smart home technologies in a health monitoring system for the elderly,” *Proc. - 21st Int. Conf. Adv. Inf. Netw. Appl. Work. AINAW’07*, vol. 1, pp. 820–825, 2007, doi: 10.1109/AINAW.2007.209.

[3] H. Pazhoumand-Dar, L. J. Armstrong, and A. K. Tripathy, “Detecting deviations from activities of daily living routines using kinect depth maps and power consumption data,” *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 4, pp. 1727–1747, Apr. 2020, doi: 10.1007/s12652-019-01447-3.

[4] I. Ekerete, N. Christopher, P. McCullagh, and J. McLaughlin, "Data Mining and Fusion Framework for In-Home Monitoring Applications," 2023.

[5] S. Kolozali, S. L. White, S. Norris, M. Fasli, and A. van Heerden, "Explainable Early Prediction of Gestational Diabetes Biomarkers by Combining Medical Background and Wearable Devices: A Pilot Study with a Cohort Group in South Africa," *IEEE J. Biomed. Heal. Informatics*, pp. 1–12, 2024, doi: 10.1109/JBHI.2024.3361505.

[6] T. Boultache, B. Achour, and M. Laghrouche, "Human activity detection from inertial data using RNN and LSTM network," *Int. J. Sens. Networks*, vol. 39, no. 3, pp. 156 – 161, 2022, doi: 10.1504/IJSNET.2022.124568.

[7] S. M. Ammal and P. S. Manoharan, "Multi-Headed Deep Learning Models to Detect Abnormality of Alzheimer's Patients," *Comput. Syst. Sci. Eng.*, vol. 44, no. 1, pp. 367 – 390, 2022, doi: 10.32604/csse.2023.025230.

[8] J. Janardhanan and S. Umamaheswari, "Exploration of Deep Learning Models for Video Based Multiple Human Activity Recognition," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, pp. 422 – 428, 2023, doi: 10.17762/ijritcc.v11i8s.7222.

[9] H. Shahverdi, M. Nabati, P. Fard Moshiri, R. Asvadi, and S. A. Ghorashi, "Enhancing CSI-Based Human Activity Recognition by Edge Detection Techniques," *Inf.*, vol. 14, no. 7, 2023, doi: 10.3390/info14070404.

[10] B. Igried, M. Alauthman, M. Alkhaldy, and A. Alsarhan, "Efficient Scanning Activity Detection in IoT Networks Using Ensemble Learning," *ACM Int. Conf. Proceeding Ser.*, p. 3625346, 2023, doi: 10.1145/3625343.3625346.

[11] F. Ribeiro, F. Fidalgo, A. Silva, J. Metrôlho, O. Santos, and R. Dionisio, "Literature review of machine-learning algorithms for pressure ulcer prevention: Challenges and opportunities," *Informatics*, vol. 8, no. 4, 2021, doi: 10.3390/informatics8040076.

[12] S. J. Mohammed, S. K. Mohsin, and M. A. Mohammed, "Healthcare Predictions Using Machine Learning and Artificial Intelligence Algorithms," *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 4, pp. 523 – 526, 2022, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85164802378&partnerID=40&md5=4a4845a0202c0406958973fd7b025206

[13] M. Jamei *et al.*, "Combined Terrestrial Evapotranspiration Index prediction using a hybrid artificial intelligence paradigm integrated with relief algorithm-based feature selection," *Comput. Electron. Agric.*, vol. 193, 2022, doi: 10.1016/j.compag.2022.106687.

[14] W. S. Monroe, F. M. Skidmore, D. G. Odaibo, and M. M. Tanik, "HihO: accelerating artificial intelligence interpretability for medical imaging in IoT applications using hierarchical occlusion: Opening the black box," *Neural Comput. Appl.*, vol. 33, no. 11, pp. 6027 – 6038, 2021, doi: 10.1007/s00521-020-05379-4.

[15] G. Szepannek and K. Lübke, "Explaining Artificial Intelligence with Care: Analyzing the Explainability of Black Box Multiclass Machine Learning Models in Forensics," *KI - Kunstl. Intelligenz*, vol. 36, no. 2, pp. 125 – 134, 2022, doi: 10.1007/s13218-022-00764-8.

[16] G. Pavlidis, "Unlocking the black box: analysing the EU artificial intelligence act's framework for explainability in AI," *Law, Innov. Technol.*, 2024, doi: 10.1080/17579961.2024.2313795.

[17] M. Garcia-Constantino *et al.*, "Probabilistic Analysis of Abnormal Behaviour Detection in Activities of Daily Living," *2019 IEEE Int. Conf. Pervasive Comput. Commun. Work. PerCom Work. 2019*, no. February, pp. 461–466, 2019, doi: 10.1109/PERCOMW.2019.8730682.

[18] Q. Ni, T. Patterson, I. Cleland, and C. Nugent, "Dynamic detection of window starting positions and its implementation within an activity recognition framework," *J. Biomed. Inform.*, vol. 62, pp. 171–180, 2016, doi: 10.1016/j.jbi.2016.07.005.

[19] I. De Falco, G. De Pietro, and G. Sannino, "Evaluation of artificial intelligence techniques for the classification of different activities of daily living and falls," *Neural Comput. Appl.*, vol. 32, no. 3, pp. 747–758, Feb. 2020, doi: 10.1007/s00521-018-03973-1.

[20] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of Explainable AI Techniques in Healthcare," *Sensors*, vol. 23, no. 2, pp. 1–19, 2023, doi: 10.3390/s23020634.

[21] T. Hulsen, "Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare," *AI*, vol. 4, no. 3, pp. 652–666, 2023, doi: 10.3390/ai4030034.

[22] N. C. Tay, T. Connie, T. S. Ong, A. B. J. Teoh, and P. S. Teh, "A Review of Abnormal Behavior Detection in Activities of Daily Living," *IEEE Access*, vol. 11, no. January, pp. 5069–5088, 2023, doi: 10.1109/ACCESS.2023.3234974.

[23] M. S. Kamal, A. Northcote, L. Chowdhury, N. Dey, R. G. Crespo, and E. Herrera-Viedma, "Alzheimer's Patient Analysis Using Image and Gene Expression Data and Explainable-AI to Present Associated Genes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–7, 2021, doi: 10.1109/TIM.2021.3107056.

[24] M. Sharma and P. Kaur, "XLAAM: explainable LSTM-based activity and anomaly monitoring in a fog environment," *J. Reliab. Intell. Environ.*, vol. 9, no. 4, pp. 463–477, 2023, doi: 10.1007/s40860-022-00185-2.

[25] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," *Data Democr. Nexus Artif. Intell. Softw. Dev. Knowl. Eng.*, pp. 83–106, 2020, doi: 10.1016/B978-0-12-818366-3.00005-8.

[26] N. Alasmari, "Mixed Methods Research: An Overview," *Int. J. Soc. Sci. Hum. Res.*, vol. 03, no. 09, pp. 44–49, 2020, doi: 10.47191/ijsshr/v3-i9-04.

[27] L. Morissette and S. Chartier, "The k-means clustering technique: General considerations and implementation in Mathematica," *Tutor. Quant. Methods Psychol.*, vol. 9, no. 1, pp. 15–24, 2013, doi: 10.20982/tqmp.09.1.p015.

[28] I. Ekerete, M. Garcia-Constantino, Y. Diaz, C. Nugent, and J. Mclaughlin, "Fusion of Unobtrusive Sensing Solutions for Sprained Ankle Rehabilitation Exercises Monitoring in Home Environments," no. August, pp. 1–11, 2021, doi: 10.20944/preprints202108.0301.v1.

[29] M. Tripathi, "Analysis of Convolutional Neural Network based Image Classification Techniques," *J. Innov. Image Process.*, vol. 3, no. 2, pp. 100–117, 2021, doi: 10.36548/jiip.2021.2.003.

[30] R. Socher, E. H. Huang, J. Pennington, A. Y. Ng, and C. D. Manning, "Dynamic pooling and unfolding recursive autoencoders for paraphrase detection," *Adv. Neural Inf. Process. Syst. 24 25th Annu. Conf. Neural Inf. Process. Syst. 2011, NIPS 2011*, no. January, 2011.

[31] M. R. Zafar and N. Khan, "Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability," *Mach. Learn. Knowl. Extr.*, vol. 3, no. 3, pp. 525–541, 2021, doi: 10.3390/make3030027.

[32] X. Xiang, H. Yu, Y. Wang, and G. Wang, "Stable local interpretable model-agnostic explanations based on a variational autoencoder," *Appl. Intell.*, vol. 53, no. 23, pp. 28226–28240, 2023, doi: 10.1007/s10489-023-04942-5.

[33] G. Amaral *et al.*, *Introduction to Algorithms for Data Mining and Machine Learning*, vol. 369, no. 1. 2013. doi: 10.1017/CBO9781107415324.004.

[34] G. Goos *et al.*, *MICCAI 2021 Part.3. Machine Learning 2*. 2021. [Online]. Available: http://www.springer.com/series/7412

[35] N. Barr Kumarakulasinghe, T. Blomberg, J. Liu, A. Saraiva Leao, and P. Papapetrou, "Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models," *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2020-July, pp. 7–12, 2020, doi: 10.1109/CBMS49503.2020.00009.