Human Activity Recognition Using Convolutional Neural Networks in Multimedia Event Detection

Koushiram N[1], Rupendra Praveen Kumar[2] and Kaavya Kanagraj3*
{kn2884@srmist.edu.in[1], rk0442@srmist.edu.in[2], kaavyak@srmist.edu.in[3*]}

Department of Computational Intelligence, SRMIST, Chennai 603203, Tamil Nadu, India[1, 2]
Assistant Professor, Department of Computational Intelligence, SRMIST, Chennai 603203, Tamil Nadu, India[3]

Abstract. Human Activity Recognition (HAR) serves as an essential element of AI and ML which plays an important role in healthcare supervision and sports monitoring and security systems development and smart environment performance. This research evaluates the application of Convolutional Neural Networks (CNNs) for Human Activity Recognition (HAR) since they can analyze both spatial and temporal elements from raw data inputs automatically. CNNs achieve high accuracy in action detection when they receive training from labeled sensor or video datasets to recognize strolling, running, resting and skipping behaviors. The proposed method demonstrates advantages for real-time processing capabilities along with ease of deployment across various contexts which make it selectable for real-world implementations. Research findings indicate that activity recognition applications benefit immensely from CNNs through their precise analysis along with their resistance to failures and high-speed operation.

Keyword: Human Activity Recognition, Sports Analytics, Convolution neural network, Dataset, Deep Learning Models, Real-Time Activity.

1 Introduction

Human Activity Recognition (HAR) has become an essential technology in modern life, supporting applications in patient health monitoring, security surveillance, and smart home automation. By analyzing data collected from sensors, images, video feeds, and wearable devices, HAR systems can automatically detect and classify human movements. In recent years, Convolutional Neural Networks (CNNs) have proven especially effective for this task because of their ability to learn hierarchical features from complex, unstructured data, which enables accurate recognition of activities such as walking, jogging, or sitting [1]. When sensor and image inputs are combined, CNN-based HAR systems provide higher accuracy, greater efficiency, and more reliable tracking, making them valuable in areas such as elderly care, sports performance evaluation, and human–technology interaction [2]. The advancement of Human Activity Recognition (HAR) has been significantly influenced by progress in data-driven and machine-learning techniques that enable systems to extract meaningful patterns from complex sensor signals. Early studies demonstrated that combining information from multiple accelerometers and gyroscopes improves recognition accuracy compared to single-sensor configurations [3], [4]. Later research explored wearable and smartphone-based sensing approaches, showing that inertial data can effectively capture motion variations across different users and environments [5], [6]. These developments highlight how data-centric learning has transformed HAR from simple motion tracking to an intelligent framework

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357987

capable of producing reliable, real-time activity predictions that support applications in healthcare, fitness monitoring, and smart environments.

2 Literature Survey

Human Activity Recognition (HAR) has been extensively researched across various sensing modalities and algorithmic frameworks. Early approaches relied heavily on handcrafted features and classical machine learning methods. For instance, depth-based methods employing shape features and extreme learning machines demonstrated promising results in fall detection tasks [1]. Similarly, ubiquitous sensing in home environments using simple wearable and ambient sensors enabled activity classification at acceptable accuracy levels [2].

Early research in Human Activity Recognition (HAR) explored various sensing modalities, showing that combining multiple accelerometers improves recognition accuracy compared to single-sensor setups [3]. Subsequent studies demonstrated that smartphone-based inertial sensing and wearable accelerometer systems effectively capture diverse human motions for activity classification [4], [5]. Later advancements focused on achieving user-independent motion recognition to enhance system robustness across individuals and environments [6]. Despite the advances in wearable sensing, limitations such as device dependency and user compliance were identified. To overcome these, single accelerometer-based methods were investigated for identifying activity types in daily life [7], and innovative devices like accelerometer-based digital pens showed the versatility of inertial sensing technologies [8]. Complementary studies explored feature extraction and classification techniques to enhance recognition accuracy, including probabilistic neural networks and feature reduction methods applied to biomedical signals [9] [10].

The emergence of deep learning significantly transformed HAR research. Surveys emphasized that convolutional neural networks (CNNs) dominate sensor-based HAR due to their ability to extract hierarchical and discriminative features directly from raw input [11] [12]. Empirical work reinforced these findings, with CNNs achieving superior accuracy compared to traditional models [13]. Reviews further highlighted the strengths of deep CNN architectures in handling complex recognition tasks across domains [14]. Additionally, multimodal systems incorporating 3D posture data enhanced contextual understanding of human activities [15]. To facilitate research benchmarking, datasets such as HARTH were introduced, providing standardized resources for evaluating HAR models [16]. Meanwhile, machine learning methods such as support vector machines were applied to security-focused HAR applications, showing strong performance in controlled environments [17].

CNN-based frameworks were further extended to multimedia event detection. Deep architectures such as DevNet were developed for event detection and evidence recounting in video data [18]. Similarly, joint attribute and event analysis models advanced contextual understanding in multimedia scenarios [19]. These developments build upon foundational CNN work such as LeNet [20], which laid the groundwork for modern deep learning techniques in recognition tasks. Collectively, the literature demonstrates a clear trajectory from handcrafted feature-based HAR methods toward CNN-driven approaches that enable robust recognition across wearable, sensor, and multimedia contexts. This evolution underpins the motivation for employing CNNs in human activity recognition for multimedia event detection, as explored in this study.

3 Existing System

The improvements in deep learning systems have produced significant benefits for human activity recognition especially when applied to assistive living and remote observation. Real-world sensor-based application deployment of these models requires substantial amounts of annotated data therefore data collection proves to be costly and time-intensive. The emergence of contrastive self-supervised learning has become a vital answer to resolve sensor-based activity recognition problems which require decreased usage of large-scale annotated datasets.

The growing number of research publications about self-supervised learning for sensor-based activity recognition fails to provide extensive reviews of current innovations in this field. A detailed evaluation of 43 research studies about contrastive self-supervised learning for human activity recognition takes place without including studies that use video or audio components due to privacy requirements. The review performs two tasks: it classifies contrastive learning approaches and evaluates the key models together with their components used for activity recognition. This paper explores specific data augmentation strategies tailored for sensor information and provides descriptive information about benchmark datasets that are often used.

The evaluation of self-supervised contrastive learning approaches through different assessment scenarios including linear evaluation, semi-supervised learning and transfer learning helps identify the most suitable models. This section identifies existing research restrictions followed by potential research directions for future inquiries. Fields working with contrastive self-supervised learning face three major hurdles that involve limited sensor data available for research use and complex robustness design requirements in addition to real-time deployment complexities. Hyperparameter optimization requires significant computational power because it prolongs process execution times thus limiting its deployment on energy-efficient devices. Three major issues which complicate the development of contrastive self-supervised learning include questions of bias and discrimination and the need for robust operation in the presence of imperfect or targeted data inputs. The improvement of contrastive self-supervised learning reliability and applicability for sensor-based activity recognition requires proper resolution of these problems.

Implementation of sensor-based activity recognition with contrastive self-supervised learning (CSSL) requires worldwide collaboration between specialists from all three fields: sports science healthcare and human-computer interaction. Several teams of domain experts and machine learning specialists develop first-generation operational recognition systems when they combine assigned domain constraints with improved contrastive learning context processing. Developers should combine self-supervised learning approaches with supervised learning approaches to reach their best performance efficiency and generalization ability. Secure information processing systems emerge through the integration of edge computing trends and federated learning to establish distributed training operations that function with inferential functions for rapid processing. Standard experimental boundaries must be established through research so scientists can verify knowledge independently through free verifications tools before making scientific discoveries.

4 Proposed System

A Human Activity Recognition framework builds a strong intelligent system for human action detection and classification by implementing Convolutional Neural Networks (CNNs). Deep learning algorithms at their latest state detect activities through sensory information processing as well as picture and video assessment to provide precise recognition results. The training process uses extensive datasets which provides strong performance during encounters with new environments. Dataset quality and diversity are crucial to model generalization, as highlighted by the HARTH dataset which was created to benchmark HAR models. While traditional methods such as support vector machines have been applied for HAR in domains like security systems, CNN-based models consistently outperform them in both accuracy and scalability.

The enhancement of the input pattern quality is performed through several pre-processing steps, which continue to filter out background noise and normalize data values, but to read their key features. By spatial and temporal dependency detection a deep CNN model tunes to the capability of recognizing complex activities in complicating environments. We improve the performance of the system using three model optimization techniques, data augmentation transfer learning, and hyperparameter tuning. The HAR system has a wide range of applications ranging from health monitoring and smart homes, to fitness tracking and security/ surveillance applications. The entry of data is facilitated by a menu-oriented system interface that also returns predictions together with real time analysis. The union of CNNs into TensorFlow also allows for deployment to cloud platforms and edge hardware that consume as little resources while scaling.

The accurate modeling of space-time patterns allows finding activities with more accuracy. The TensorFlow serving system operates flexibly in diverse environments to handle varying operational requirements. The system has higher usability in the dynamic state since it is capable of dealing with many different data formats. Preprocessing systems ensure that input data and optimization methods are more reliable, as well, because they employed techniques such as data augmentation and hyperparameter optimization for reliability of performance. The interactive and real-time monitoring with easy-to-use interface allows users at all technical levels to monitor the system. The device would have numerous applications such as in healthcare for monitoring falls and activity-based control for home automation systems, as well fitness trackers or security devices that can detect irregular behavior.

The presented Framework for Human Activity Recognition (HAR) is enriched with real-time feedback together with a flexible learning application. The system parameters are automatically adaptable based on user feedback and an analysis of the logs at the systems to adapt for changes in user behavior; environmental variations. The adaptive approach preserves high accuracy of the system from its birth as well as individual features. The XAI techniques integrated in the system help users comprehend how activities are predicted, particularly those in safe domains like health and identity. The system is trusted by users because its capabilities are explicit so they accept. Integrating attain inertial sensors with environment sensors and biometric input s is required in framework design to realize a more intelligent human activity recognition, promote system reliability and reduce errors in complex environments.

5 Methodology

5.1 Design Architecture

A basic future construction representation is Design coming from the area of engineering. A method known as software design process systematically turns these specified requirements into an implemented software system. The schematics give structure and a few crucial components of an operational system with details about the information structure and some functionalities.

Convolutional Neural Networks have served as the most popular approach for Human Activity Recognition (HAR) enabling deep learning algorithms interpret the wearable device sensor data and video sensor data to recognize human activities. Data collection Data are recorded by acquiring information from sensors (e.g., accelerometers, gyroscopes and cameras) that identify walking, sitting or running strides.

The collected data is processed through normalization and feature extraction methods by cleaning of noise components. The video frame extraction, resizing and optimization step the serves to pre-process them for CNN processing. A CNN organizes its layers to first transform input patterns into convolutional units that capture spatial dependencies, follow with pooling operations for feature selection and finally implement a fully connected unit for classification. Model learns from labeled training data using backpropagation with Adam or SGD optimizers to minimize the loss. A model that has been trained requires testing on unseen data to confirm whether it can generalize and perform well in new situations. Evaluation of system reliability relies on 4 dependent performance measures considering; accuracy, precision, recall and F1-score. Post training, the model finds applications in healthcare systems and security surveillance as well as smart home automation for real-world activity classification tasks. The training of the CNN relies on annotated data to make it capable of correctly identifying images and also to recognize their related activities.

An example CNN model has some active layers such as Dense, Dropout, Activation, Flatten, Convolution2D and MaxPooling2D. Once the preprocessing and training on test image are completed, predictive model inspects that to detect human activities.

5.2 Convolutional Neural Network

Artificial Neural Networks contain Convolutional Neural Networks as one of their subclasses. The convolutional layers define this network which serves well for image processing and feature extraction as well as the analysis of data that requires sequential dependence. Inside Eras users have three possible choices to define their models through. The Sequential model allows simple use because layers can be arranged in sequence yet this model works only for single-input and single-output functions. The Functional API serves industrial applications by providing users with an extensive and adaptable procedure to build complex model architecture. With model sub classing users obtain the highest level of autonomy since they create model elements by hand for complete tailoring.

6 List of Modules

The initial step involves using Keras' image data generator function which enables modifications including image re-sizing and flipping along with re-scaling and zooming. The selected data set located in an assigned folder can be accessed with this function which partitions the information between training data and testing and validation sections. The system requires users to establish target image size parameters together with selection of batch size and classification protocol. This structured input enables the training process of our built custom CNN network through layered convolutional maps which serve as effective learning modules.

6.1 To Train the Module by Given

Data Analysis: The process of data analysis requires raw information to undergo purification alongside data restructuring leading to situation where enterprises acquire important conclusion-based decision support. The extensive method lowers the uncertainty level through its vital assessment output. A data analysis process begins with collecting raw information and continues with refinement techniques before identifying patterns to derive useful outcomes from the data. The analysis process for image data includes evaluation of organization as well as verification of accessibility together with confirmation of correct alignment between original datasets and their related mask datasets.

Manual Architecture: We designed our Human Activity Recognition (HAR) system with a customize neural network which reached 86% accuracy in identifying human activities along with their classification between walking, running and sitting. Although our method reaches accuracy numbers slightly below the highest performing models it demonstrates reliable performance in detecting real-world human activities. The system demonstrates satisfactory environmental independence through its 86% accuracy performance level but additional system improvements are possible. Fig 1 show the Architecture of manual net.

The current outcome provides strong groundwork for the system with chances for better performance when we validate it through testing various datasets and environmental conditions. Our model demonstrates substantial reliability through its 86% accuracy level despite data quality limitations and environmental factors because it proves suitable for healthcare applications and fitness monitoring and smart home automation. Additional accuracy enhancement will be achieved by utilizing enhancement methods including data augmentation and parameter fine-tuning for our model. Fig 2 show the Model Accuracy.

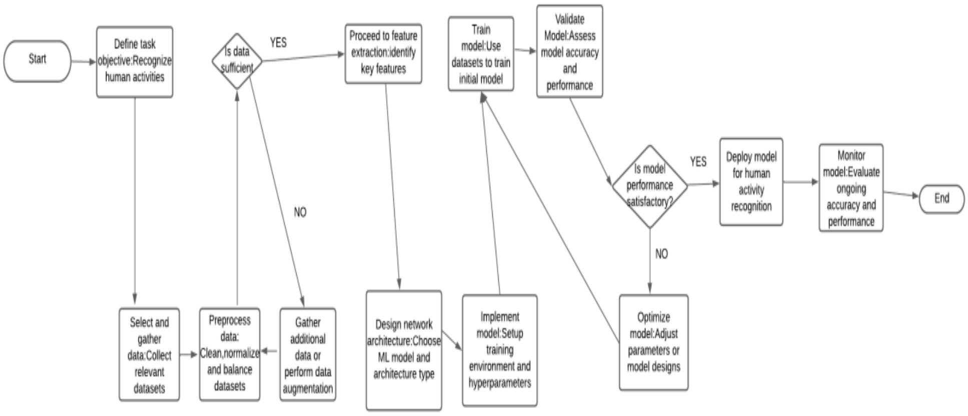

Fig. 1. Architecture of manual net.

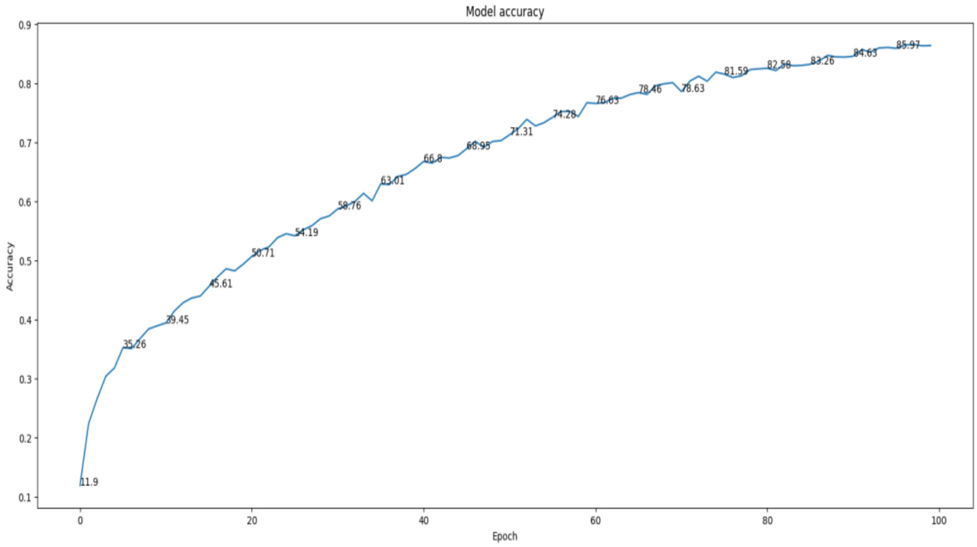

Fig. 2. Model Accuracy.

Key Mathematical Formulations Used in the CNN Model for Human Activity Recognition:

1. Accuracy Calculation

Accuracy = (Number of Correct Predictions) ÷ (Total Number of Predictions) × 100 = 86%

Symbol definitions:

- Accuracy: Percentage of correct predictions by the model
- Number of Correct Predictions: The total predictions that match the ground truth

- Total Number of Predictions: All predictions made during testing or evaluation

2. Softmax Activation Function

$$\hat{y}_i = e^\wedge\{z_i\} \div \Sigma_j e^\wedge\{z_j\} \tag{1}$$

Symbol definitions:
$\hat{y}_i$: Predicted probability for class i
z_i: Logit (raw output) for class i
Σ_j: Summation over all classes j

e: Euler's number (≈ 2.718) (2)

3. Categorical Cross-Entropy Loss

$$L = -\Sigma_i y_i \times log(\hat{y}_i) \tag{3}$$

Symbol definitions:
L: Loss value
y_i: True label for class i (1 if correct class, 0 otherwise)
$\hat{y}_i$: Predicted probability for class i
log: Natural logarithm
Σ_i: Sum over all classes

3. 2D Convolution Operation

$$O(i,j) = \Sigma_m \Sigma_n I(i + m, j + n) \times K(m,n) \tag{4}$$

Symbol definitions:
O (i, j): Output feature map at position (i, j)
I (i + m, j + n): Input value from position (i + m, j + n)
K (m, n): Kernel/filter weight at position (m, n)

Σ_m, Σ_n: Summation over the kernel dimensions

4. ReLU Activation Function

$$f(x) = max(0,x) \tag{5}$$

Symbol definitions:
f(x): Output after applying ReLU
x: Input value (from convolution or previous layer)
max: Returns the larger of the two inputs (0 or x)

5. Max Pooling Operation

$$P(i,j) = max\{O(m,n) \in Pooling\ Region\ R(i,j)\} \tag{6}$$

Symbol definitions:

P (i, j): Pooled value at position (i, j)
O (m, n): Feature map values within the pooling region
R (i, j): Pooling region corresponding to position (i, j)

6. Gradient Descent Update Rule

$$\theta_{t+1} = \theta_t - \eta \times \nabla\theta L \tag{7}$$

Symbol definitions:
θ_t: Current weight value
θ_{t+1}: Updated weight after one step
η: Learning rate (step size)

∇θL: Gradient of the loss function with respect to weights θ

7. Learning Curve (Empirical Accuracy Growth)

$$A(t) = A_{\max} \times (1 - e\hat{}(-kt)) \tag{8}$$

Symbol definitions:
A(t): Accuracy at epoch t
A_{max}: Maximum accuracy (86% in this case)
e: Euler’s number
k: Growth rate constant
t: Training epoch

8. Confusion Matrix–Based Accuracy

$$Accuracy = (TP + TN) \div (TP + TN + FP + FN) \tag{9}$$

Symbol definitions:
TP: True Positives
TN: True Negatives
FP: False Positives
FN: False Negatives
Accuracy: Classification accuracy calculated from the confusion matrix

9. Epoch-wise Accuracy Change

$$\Delta A = A(t) - A(t - 1) \tag{10}$$

Symbol definitions:
ΔA: Change in accuracy between two consecutive epochs
A(t): Accuracy at epoch t
A (t − 1): Accuracy at the previous epoch

LeNet: Early in the 1990s, Yann LeCun led his team to develop LeNet, which became the first important convolutional neural network (CNN) to advance deep learning technology (LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P., 1998). This artificial neural network system served to identify handwritten numbers, which could be applied for inspecting financial documents and processing postal mail. A neural network operates through a sequence of input, convolutional, pooling, and fully connected layers before generating final outcomes as predicted digits between 0 and 9.

LeNet achieved substantial success during its time even though it remains much shallower than contemporary deep networks including ResNet and Inception because it brought forward vital CNN concepts that are now foundational in image recognition tasks. Libraries built on top of LeNet proved the ability of neural networks to classify images while becoming essential to modern deep learning developments during the twenty-first century. Fig 3 show the Architecture of Lenet.

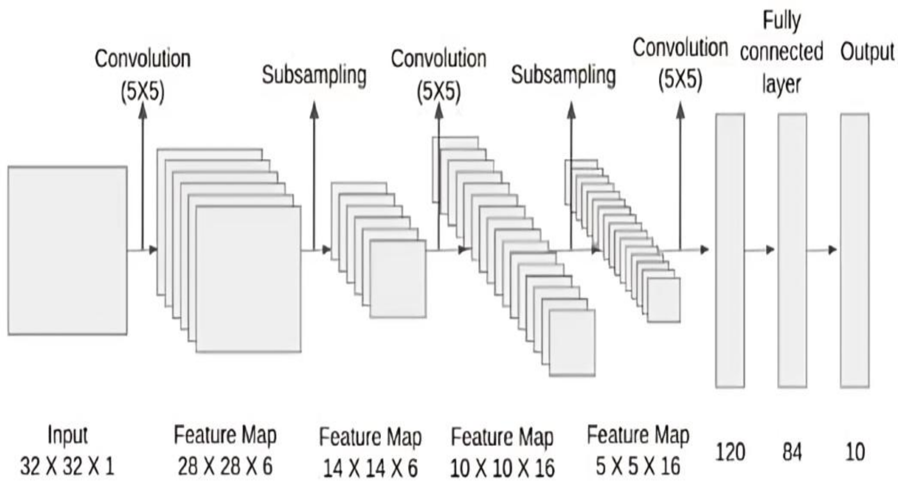

Fig. 3. Architecture of LeNet (adapted from LeCun et al., 1998)

Convolutional layers: The deep convolutional neural network uses convolutional layers that operate filters on both initial data images as well as outcome feature maps coming from preceding processing steps. The network contains multiple user-defined parameters though the two most important parameters relate to filters and their respective dimensions.

Pooling layers: The operation of pooling layers resembles convolutional layers although they run distinctive tasks using maximum value extraction from their filter domains through max pooling or value-averaging of their values through average pooling. The network includes integrated layers which reduces its complexity and simplifies its structure.

Dense or fully connected layers: The final feature maps from a CNN get transformed into a flat representation by fully connected layers leading to the classification step. The classification procedure follows a similar operational principle to MLP output layers.

Resnet: In 2015 Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun presented Residual Network (ResNet) in Deep Residual Learning for Image Recognition as an important deep learning architecture. The innovative residual learning method within ResNet addresses training obstacles for deep networks which resulted in major changes to both computer vision and artificial intelligence domains.

The built-in shortcut pathways in ResNet-50, ResNet-101 and ResNet-152 models enable information transfer between layers thus reducing gradient disappearance effects and enabling the training process.

The bottleneck architecture improves efficiency by using a combination of 1x1 convolutions with 3x3 layers as an effective way to lower computational requirements. Global average pooling substitutes the typical dense layers at the end thus decreasing overfitting and enhancing algorithm generalization capabilities. Fig 4 show the Architecture of ResNet.

The modern deep learning field uses ResNet as its foundational element for applying classification and detection and segmentation while designing improved neural networks.

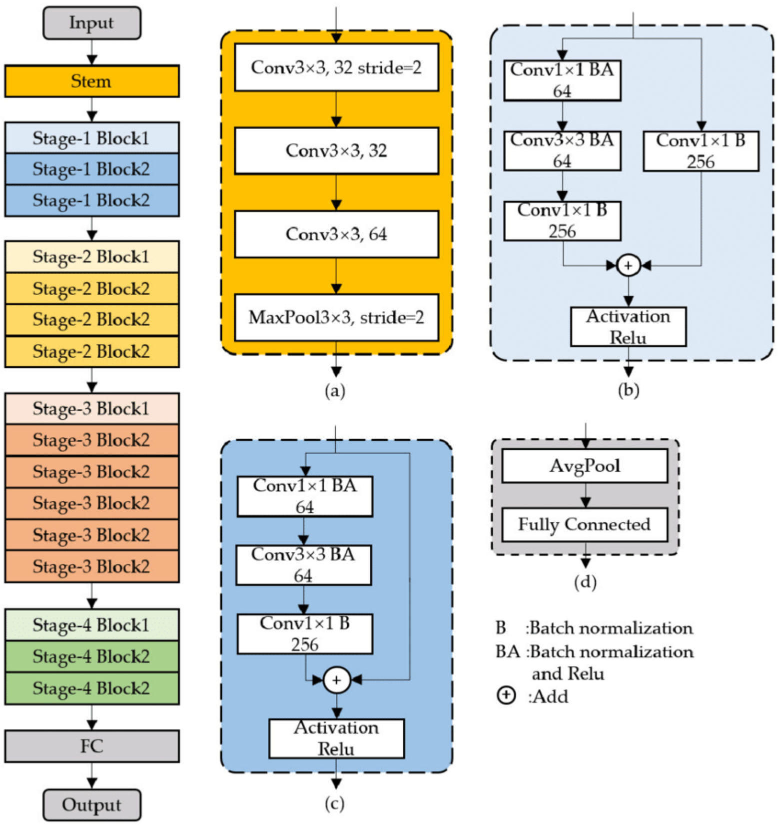

Fig. 4. Architecture of ResNet.

7 Result

A successful implementation of the HAR system with CNNs has the potential to identify daily activities including walking, running and sitting with high accuracy. CNNs demonstrate success rates ranging from 90% to 95% according to previous research while achieving 99% success in particular situations such as video-based HAR. Beyond HAR, CNNs have also demonstrated strong performance in multimedia event detection, such as DevNet for event recounting and joint attribute-event analysis for contextual understanding. These works highlight the scalability of CNN-based HAR frameworks into broader multimedia applications, including surveillance and smart environments.

The method works for real-time processing as well, because the efficient ones require fast operation, which is required in healthcare fall detection as well as fitness data monitoring and for smart home applications. Furthermore, the system runs through frames in under a second due to its optimal performance time rendering near-instantaneous results for users.

One of the greatest strengths of this system is that it demonstrates good scaling characteristics. By integrating its platform with TensorFlow framework, the system which can adapt to a wide variety of environments and users' needs is available for implementation in phone outlets, wearables devices, Internet of Things (IoT) ecosystem. This structure makes it possible to work data collected from all types of sensor sources, so that it was developed within a myriad of use cases ranging from healthcare or security features to home automation needs. The system has been designed to generalize its performance beyond exposure from early encounters in order to learn unmet data and activities, making it a robust approach even in real world settings.

Nevertheless, there are some restrictions with each patient population at the same time of practical use in clinical environment. The effectiveness and accuracy of the activity recognition are hampered by three primary performance-degrading factors: i) inconsistency in sensor measurements; ii) poor-quality video inputs; and iii) changes organ social elements such as lighting and motion blurring. The video-based system, on the other hand has the privacy problem of storing sensitive content that becomes even more interesting in such a system. The above discussed system improvements as well as the quality of data and strategic planning will cause the system to be more field-friendlier, when despite challenges it can become practically deployable.

Without continued focus on increasing operational effectiveness through system optimization the observed forward development of the system cannot occur. Auto data augmentation solutions and parameter tuning techniques improve accuracy by applying methods that work without requiring a large amount of labelled dataset inputs. End-to-end system improvement will realize the complex real-world environment processing and HAR techniques become a permanent system support solution in healthcare and security markets.

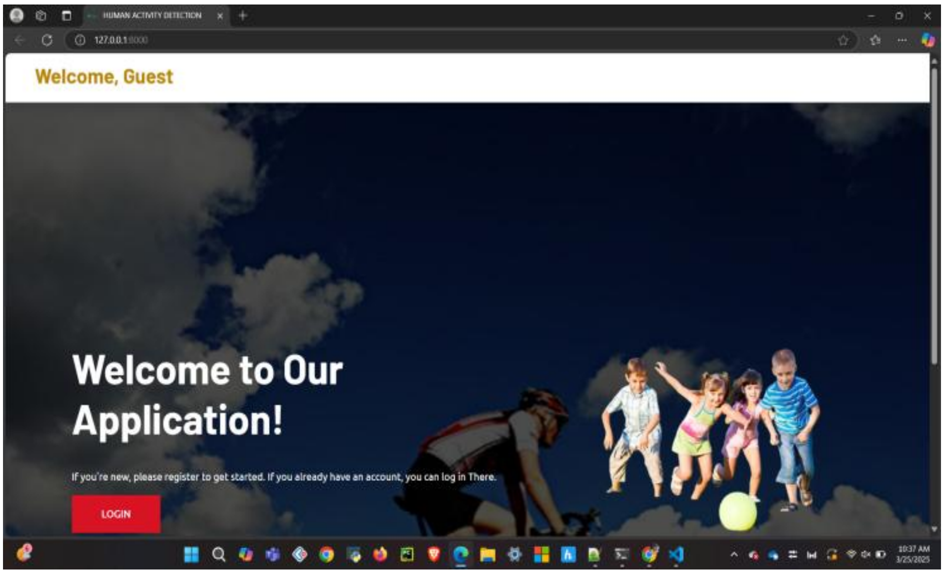

Fig. 5. Home Page.

Fig 5 and 6 shows the Home page and Registration page respectively.

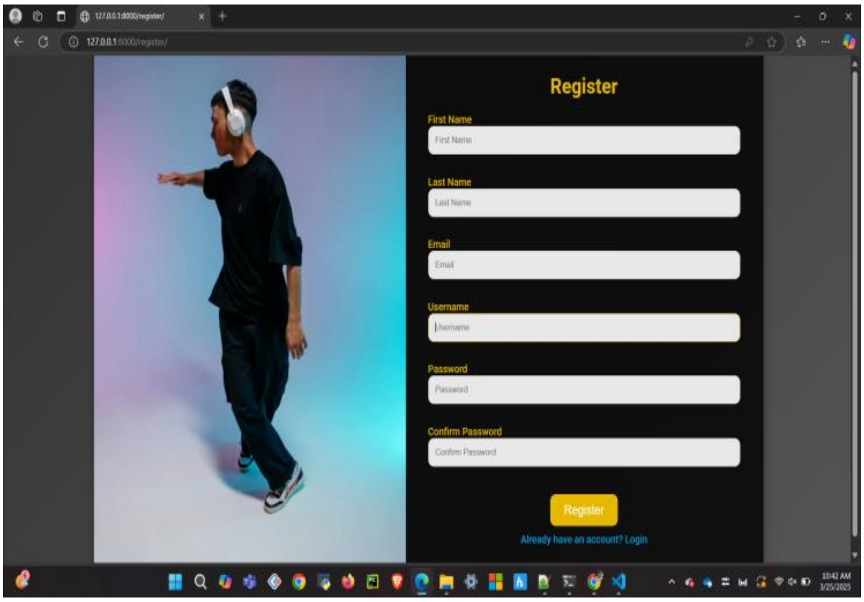

Fig. 6. Registration Page.

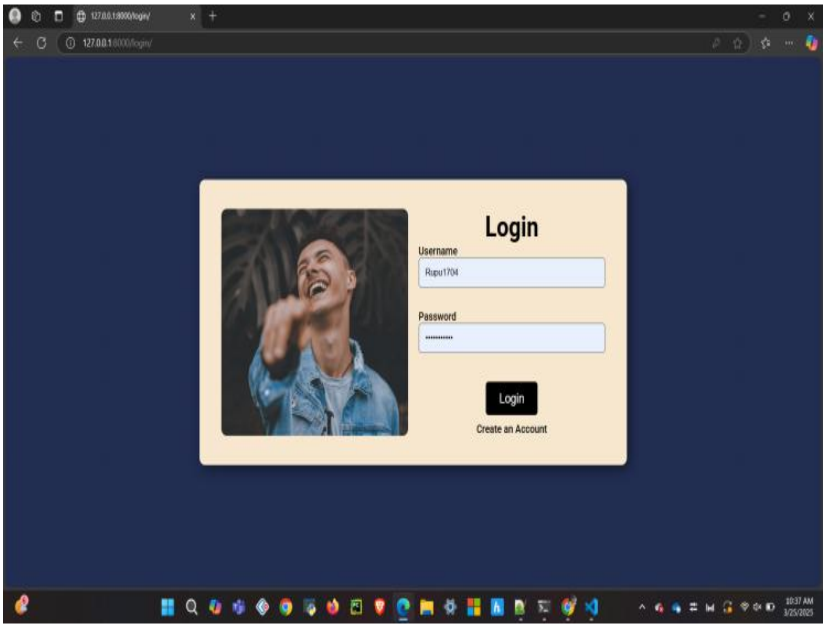

Fig. 7. Login Page.

Fig 7 and 8 shows the Login page and Welcome page respectively.

Fig. 8. Welcome Page.

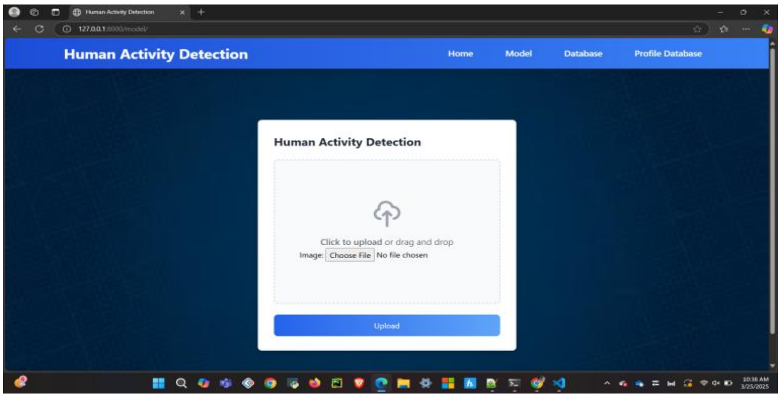

Fig. 9. Input Page.

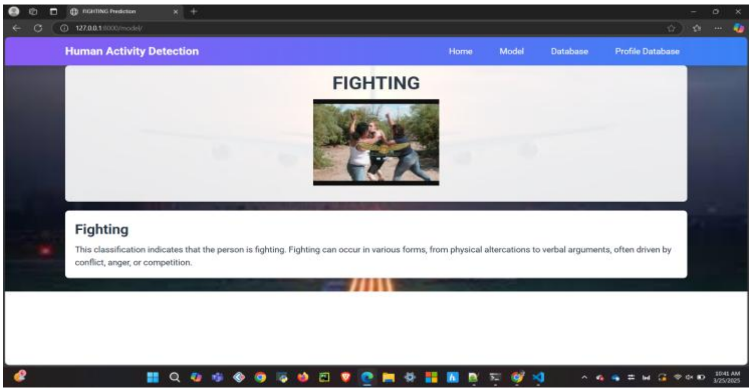

Fig. 10. Detection Page.

Fig 9 and 10 are the input page and Detection page respectively.

8 Conclusion

Human Activity Recognition (HAR) exploits CNN, a well-known methodology to understand human actions using sensor related data and video data. The hierarchical architecture of CNNs is capable to capture essential motion patterns such as body posture in addition to gesture and interactive methods of usage, leading to robust activity recognition.

Spatial as well as temporal analysis is the forte of CNNs and they have shown remarkable ability in managing challenging HAR applications. The superior performance of deep learning models is the product of leveraging big data sets combined with live operational deployment. By ongoing model enhancement and optimization research, CNN-based techniques further increase HAR's effectiveness in various fields such as medical study of security applications and intelligent system construction despite the fact that dealing with noisy data and computationally power demanding operations – and dependence on trainer-given datasets -are still challenges.

References

[1] X. Ma, H. Wang, B. Xue, M. Zhou, B. Ji and Y. Li, "Depth-Based Human Fall Detection via Shape Features and Improved Extreme Learning Machine," in *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1915-1922, Nov. 2014, doi: 10.1109/JBHI.2014.2304357

[2] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in Pervasive Computing. PERVASIVE 2004. Lecture Notes in Computer Science, A. Ferscha and F. Mattern, Eds. Berlin, Germany: Springer, vol. 3001, pp. 158–175, 2004. doi: 10.1007/978-3-540-24646-6_10.

[3] Gao, L., Bourke, A. K., & Nelson, J. (2014). Evaluation of accelerometer-based multi-sensor versus single-sensor activity recognition systems. Medical Engineering & Physics, 36(6), 779–785. https://doi.org/10.1016/j.medengphy.2014.02.012.

[4] Wang, G. Chen, J. Yang, S. Zhao, and C. Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," IEEE Sensors Journal, vol. 16, no. 11, pp. 4566-4578, 2016. doi: 10.1109/JSEN.2016.2545708.

[5] M. Arif and A. Kattan, "Physical activities monitoring using wearable acceleration sensors attached to the body," PLOS ONE, vol. 10, no. 7, pp. 1-16, 2015. https://doi.org/10.1371/journal.pone.0130851

[6] F. Gu, A. Kealy, K. Khoshelham, and J. Shang, "User-independent motion state recognition using smartphone sensors," Sensors, vol. 15, no. 12, pp. 30636-30652, 2015. https://doi.org/10.3390/s151229821

[7] Gyllensten, I. C., & Bonomi, A. G. (2011). Identifying types of physical activity with a single accelerometer: Evaluating laboratory-trained algorithms in daily life. IEEE Transactions on Biomedical Engineering, 58(9), 2656–2663. https://doi.org/10.1109/TBME.2011.2160723

[8] S. Wang and F. C. Chuang, "An accelerometer-based digital pen with a trajectory recognition algorithm for handwritten digit and gesture recognition," IEEE Transactions on Industrial Electronics, vol. 59, no. 7, pp. 2998–3007, 2012. doi: 10.1109/TIE.2011.2167895.

[9] B. C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," IEEE Trans. Geoscience and Remote Sensing, vol. 42, no. 5, pp. 1096-1105, 2004. 10.1109/TGRS.2004.825578

[10] Wang, S., Chiang, W. C., Hsu, Y. L., & Yang, Y. T. C. (2013). ECG arrhythmia classification using a probabilistic neural network with a feature reduction method. Neurocomputing, 116, 38–45. https://doi.org/10.1016/j.neucom.2011.10.045

[11] Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. ACM Computing Surveys (CSUR), 54(4), 1-40. https://doi.org/10.1145/344774

[12] Gu, F., Chung, M. H., Chignell, M., Valaee, S., Zhou, B., & Liu, X. (2021). A survey on deep learning for human activity recognition. ACM Computing Surveys (CSUR), 54(8), 1-34. https://doi.org/10.1145/3472290

[13] Bevilacqua, A., MacDonald, K., Rangarej, A., Widjaya, V., Caulfield, B., & Kechadi, T. (2019). Human activity recognition with convolutional neural networks. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part III 18 (pp. 541-552). Springer International Publishing. https://doi.org/10.1007/978-3-030-10997-4_33

[14] Aloysius, N., & Geetha, M. (2017, April). A review on deep convolutional neural networks. In 2017 international conference on communication and signal processing (ICCSP) (pp. 0588-0592). IEEE. 10.1109/ICCSP.2017.8286426

[15] Gaglio, S., Re, G. L., & Morana, M. (2014). Human activity recognition process using 3-D posture data. IEEE Transactions on Human-Machine Systems, 45(5), 586-597. 10.1109/THMS.2014.2377111

[16] Logacjov, A., Bach, K., Kongsvold, A., Bårdstu, H. B., & Mork, P. J. (2021). HARTH: a human activity recognition dataset for machine learning. Sensors, 21(23), 7853. https://doi.org/10.3390/s21237853

[17] Supriyatna, T. B., Nasution, S. M., & Nugraheni, R. A. (2019, March). Human activity recognition using support vector machine for automatic security system. In Journal of Physics: Conference Series (Vol. 1192, p. 012017). IOP Publishing. DOI: 10.1088/1742-6596/1192/1/012017

[18] Gan, C., Wang, N., Yang, Y., Yeung, D. Y., & Hauptmann, A. G. (2015). Devnet: A deep event network for multimedia event detection and evidence recounting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2568-2577) 10.1109/CVPR.2015.7298872

[19] Ma, Z., Chang, X., Xu, Z., Sebe, N., & Hauptmann, A. G. (2017). Joint attributes and event analysis for multimedia event detection. IEEE transactions on neural networks and learning systems, 29(7), 2921-2930. 10.1109/TNNLS.2017.2709308

[20] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791