Spak-Net: A Lightweight Convolutional Neural Network for Activity Recognition with Wearable Inertial Sensors

Shaida Muhammad
Department of Computing
National University of Sciences & Technology (NUST)
Islamabad, Pakistan
smuhammad.mscs19seecs@seecs.edu.pk

Hamza Ali Imran
Department of Computing
National University of Sciences & Technology (NUST)
Islamabad, Pakistan
himran.mscs18seecs@seecs.edu.pk

Wasim Zaman
Department of Electrical, Electronic and Computer Engineering
University of Ulsan
Ulsan, South Korea
wasim94@mail.ulsan.ac.kr

Kiran Hamza
Department of Computing
National University of Sciences & Technology (NUST)
Islamabad, Pakistan
kmehmood.mscs19seecs@seecs.edu.pk

***Abstract*—The prevailing Internet of Health and Medical Things (IoHMT) strategy emphasizes preventing disease onset by continuously monitoring individuals' physical activities. This approach makes Human Activity Recognition (HAR) and Behavior Analysis crucial areas of research in the field of IoHMT. A specific subset of HAR, known as Sports Activity Recognition (SAcR), aims to detect and classify sports movements. Three primary methodologies are used to monitor such activities: computer vision, environmental sensors, and wearable sensors. Among these, wearable sensors emerge as the most viable solution, based on an assessment of the advantages and limitations of each technology. This paper introduces Spak-Net, a model characterized by its efficient use of addition layers and containing only 398 trainable parameters. Leveraging inertial sensor data from wearable devices, the model demonstrated exceptional performance, achieving 98.93% ± 0.9% (95% CI) accuracy on the sports activity recognition IM-Sporting Behavior dataset and 96.25% ± 1.2% (95% CI) accuracy on the human activity recognition WISDM-11 dataset making it the most efficient neural network for achieving this much performance on WISDM11 dataset to the best of our knowledge.**

***Index Terms*—Human Activity Recognition, Sports Activity Recognition, Internet of Health and Medical Things, Human Behavioral Analysis, Inertial Sensors**

I. Introduction

Human Activity Recognition (HAR) has given rise to a specialized domain called Sports Activity Recognition (SAcR), which aims to identify specific sports activities [1]. This field's significance is underscored by the need for strategic sensor placement on the human body to capture accurate and relevant data from wearable inertial devices. Recent advances in smart textiles have enabled seamless sensor integration into clothing, enhancing user comfort and monitoring capabilities. To achieve precise and reliable recognition, careful sensor placement on the body is crucial. Wearable inertial devices can capture valuable motion data when strategically positioned, enhancing the accuracy of activity recognition systems.

ISAcR is particularly valuable in optimizing athletic performance and injury prevention. Our research leverages the wearable sensor dataset provided by Jalal et al. [2].

As SAcR continues to evolve, it presents considerable potential for advancing sports science and related fields in Pakistan and beyond. SAcR technology also has promising applications in monitoring physical health and rehabilitation for injured individuals. SAcR-based systems can assist physiotherapists in tracking recovery progress in patients with musculoskeletal injuries, offering objective insights that complement traditional clinical observations. In the context of sports training, SAcR enables real-time feedback for athletes and coaches which can be instrumental in refining techniques and preventing injuries. This technology is adaptable to various sports activities, enhancing its utility in sports science and rehabilitation.

Deep Neural Networks have become increasingly prevalent in HAR and SAcR. However, their complexity renders them unsuitable for edge device deployment, especially in real-time domains. To address this challenge, we propose 'Spak-Net', a novel lightweight model with only 398 trainable parameters. The name Spak-Net is derived from the Pashto word "Spak," which means "light" or "small," reflecting the reduced number of parameters. We built the model as an improved version of the previous model 'Mukhtasir-Khail-Net' [3]. Notably, Spak-Net boasts a reduction in complexity compared to its predecessor Mukhtasir-Khail-Net with better accuracy. Spak-Net is best suited for real-time application in edge devices.

Contributions of the article:

979-8-3315-1533-1/25/$31.00 ©2025 IEEE

This is the preprint version. Copyright belongs to IEEE. Kindly cite the published version: https://ieeexplore.ieee.org/document/11034469

- We presented Spak-Net which has carefully selected layers and has only 398 trainable parameters. We are making 1.6 times less complex than its predecessor.
- Spak-Net has been evaluated on IM-SportingBehaviors dataset [2]. The average accuracy achieved is 98.93%. The F1-score, Recall, and Precision all reached a remarkable 98.93%, making it outperform its predecessor's performance. Details are presented in section IV-B.
- Spak-Net is also evaluated on the WISDM 11 dataset [4] and it has achieved 96.25% average Accuracy, F1, Recall and Precision. This makes it the most efficient HAR model presented to date among comparable models. The performance details on HAR are reported in section IV-A.

The remaining of the article has following sections. Section II discusses the literature, and section III shares details of the proposed model, the dataset used, segmentation, and training process. Section IV contains details of results achieved with a comparison with previous work. Section V concludes the article.

II. Related Work

Sports Activity Recognition (SAcR) falls within the broader area of Human Activity Recognition (HAR). The field of SAcR with inertial sensor data has seen substantial advancements with the integration of wearable sensors. To provide a comprehensive foundation, this review includes studies primarily focused on HAR through inertial sensor data.

In one significant work, Imran et al. [5] introduced an edge-optimized neural network specifically for SAcR, built with a minimal architecture of convolutional layers and containing only 1,251 parameters. Another edge-friendly model [6] contained 2,031 trainable parameters which is designed to classify six distinct daily activities with high accuracy. In a related effort, Hamza et al. [7] proposed a model inspired by the ResNeXT architecture, achieving a 96.62% accuracy for classifying daily activities by applying deep convolutional layers. Expanding upon convolutional approaches, Imran et al. [8] created a BiGRU-CNN hybrid model, achieving an impressive 97.2% accuracy using smartphone-based inertial data.

Some studies have explored models inspired by dense neural networks. For instance, Mehmood et al. [9] introduced a DenseNet-based architecture with approximately 6 million parameters, achieving an accuracy of 93% on daily activity recognition tasks using mobile inertial sensors. Another model, HHARNet, was presented by Imran et al. [10]. This network utilized dense connections between inception-like modules of various kernel sizes, reaching 95% accuracy in classifying daily activities from smartphone inertial data.

Beyond daily activity recognition, other studies have demonstrated the diverse potential of wearable inertial sensors across fields such as healthcare and sports. Sabatini et al. [11] and Riaz et al. [12] leveraged motion data to analyze physical attributes like gait, height, and age, showing the usefulness of wearable technology in healthcare monitoring outside conventional clinical environments. Wearable sensors also enable applications in emotional state analysis, cognitive function monitoring, and fall detection [13]–[15].

More complex architectures have also been tested for HAR in sports contexts. Qi et al. [16] developed a hybrid model, GPARMF, incorporating a multi-layer approach that combines a one-class SVM, Hidden Markov Model, and a neural network for gym action recognition. The model achieved an average F-measure of 90.78%. Lu et al. presented a HAR model using accelerometers and Gradient Boosting Decision Trees (GBDT), which reached 93% accuracy on the DALIAC dataset. Other studies by Xia et al. [17] employed LSTM and CNN layers for enhanced sequence processing, achieving over 95% accuracy on multiple datasets including the popular UCI-HAR dataset.

Recently, efficient architectures have gained attention for their capacity to deliver high accuracy with low computational requirements. Muhammad et al. [1] proposed a CNN model with residual connections to process accelerometer and gyroscope data without extensive feature extraction. Tested on the WISDM dataset, the model achieved an accuracy of 98.32% using only 38,342 parameters. Similarly, our previous Mukhtasir-Khail-Net [3] demonstrated the potential of minimalistic architectures for SAcR with only 651 parameters and an accuracy of 98.87% on the IM-Sporting Behaviors dataset. Generisch-Net, introduced by Hamza et al. [18] used BiGRU and CNN layers, achieving 96.97% accuracy for classifying daily activities, and demonstrates the adaptability of generalized frameworks for diverse HAR tasks.

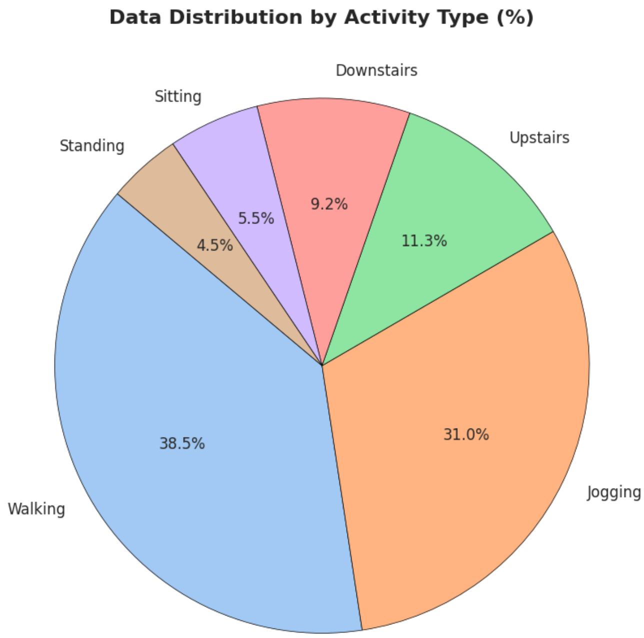
As research evolves, interest is growing in developing low-cost, scalable computational solutions for HAR. For example, Beowulf clusters constructed from Raspberry Pi devices simulate supercomputing capabilities, which can handle intensive tasks in wearable sensor data processing [19]. The use of parallel computing setups with master-slave node configurations has been shown to efficiently execute complex algorithms, such as Monte Carlo simulations, supporting HAR tasks on resource-limited devices [20].

III. Methodology

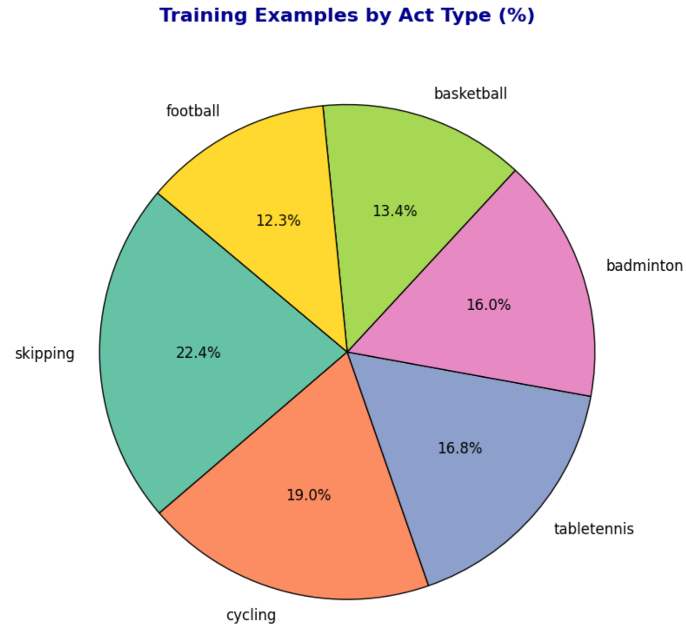
We use Spak-Net for human activity recognition and sports activity recognition tasks. The data used consists of readings from three accelerometers, making it 9-dimensional. We have used two sensor data i.e. features, the ones with no missing values making the input equal to 6D. The data is time series data therefore, we first created its segments. The size of the segmentation window is an important parameter the details are present in section III-C. The segments are then used to train the presented model. It is to be noted that we have not done any preprocessing to keep the solution less complex and entirely a deep-learning one. The raw sensor data is being used. The details of the dataset, presented model, and segmentation are presented in this section.

This is the preprint version. Cite published paper: https://ieeexplore.ieee.org/document/11034469

This is the preprint version. Copyright belongs to IEEE. Kindly cite the published version: https://ieeexplore.ieee.org/document/11034469



(a) Activity dataset: WISDM-11



(b) Sports dataset: IM-SportingBehaviors

Fig. 1. Class distribution for activity and sports datasets

A. Datasets

Researchers from Air University Pakistan created and shared the IM-SportingBehaviours 1 dataset [2]. This dataset captures key aspects of human mobility by attaching triaxial accelerometers to various parts of the body, including the knee, wrist, and below-the-neck region. The dataset includes motion data for six different sports: cycling, badminton, skipping, basketball, football, and table tennis. The dataset included both professional and amateur athletes aged 20 to 30, weighing between 60 and 100 kilograms. Figure 1 shows the class distribution in the dataset. The dataset has an imbalance, with the 'Skipping' class accounting for the majority (22.4%). Cycling accounts for 19% of the dataset, followed by Table Tennis (16.8%), Badminton (16%), Basketball (13.4%), and Football (12.3%).

The distribution of IM-Sporting Behavioral is provided in figure 1 which contains different activities.

B. Spak-Net

Spak-Net processes input data $X \in \mathbb{R}^{256\times 6}$ through two parallel 1D convolutional layers with kernel sizes 3 and 5, generating feature maps F_1 and F_2. These pass through another set of convolution layers, producing F_3 and F_4, which are fused using element-wise addition: $F_{\text{sum}} = F_3 + F_4$. A Global Average Pooling (GAP) layer then reduces the feature maps by averaging across the time dimension: $z = \frac{1}{T}\sum_{t=1}^{T} F_{\text{sum},t}$. Finally, a fully connected dense layer applies a transformation $\hat{y} = \sigma(W_d z + b_d)$ to generate predictions.

The model 2 processes raw input data through a two-path convolutional strategy. The input data shaped as $batch \times 256 \times 6$, is passed through two separate convolutional layers with differing kernel sizes. One path applies a 1×3 convolutional kernel, while the other uses a 1×5 kernel. Each of these convolutional layers outputs 6 feature maps, maintaining consistency with the input channel size. The outputs of these layers are combined through an Add operation so the feature dimensions remain unchanged.

Subsequently, the combined output undergoes another set of parallel convolutional layers repeating the 1×3 and 1×5 convolutions. The results of these layers are again merged using an Add layer.

The aggregated features are then processed using Global Average Pooling 1D which reduces the spatial dimensions while preserving the number of feature maps. This operation ensures the feature dimensions are compact and ready for classification. Finally, the model concludes with a Dense layer containing a 6×6 kernel followed by a softmax activation function to output the class probabilities for the given dataset.

This architecture ensures efficient feature extraction and classification while maintaining computational efficiency through the use of additive feature fusion. We further explain the benefits and motivations of replacing the concatenation layer with an addition layer in detail.

The addition-based fusion mechanism was selected over concatenation to maintain parameter efficiency while preserving discriminative features. Kernel sizes (3,5) were chosen based on empirical experiments demonstrating their ability to

This is the preprint version. Copyright belongs to IEEE. Kindly cite the published version: https://ieeexplore.ieee.org/document/11034469

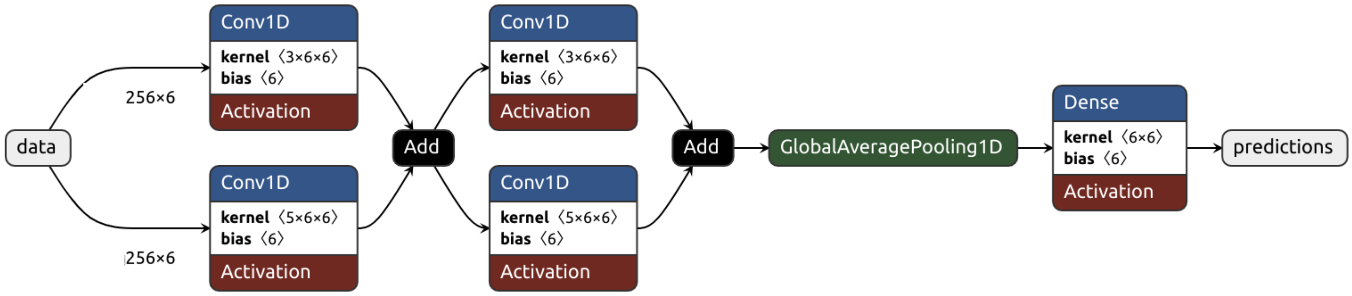


Fig. 2. Spak-Net: The Proposed Neural Network

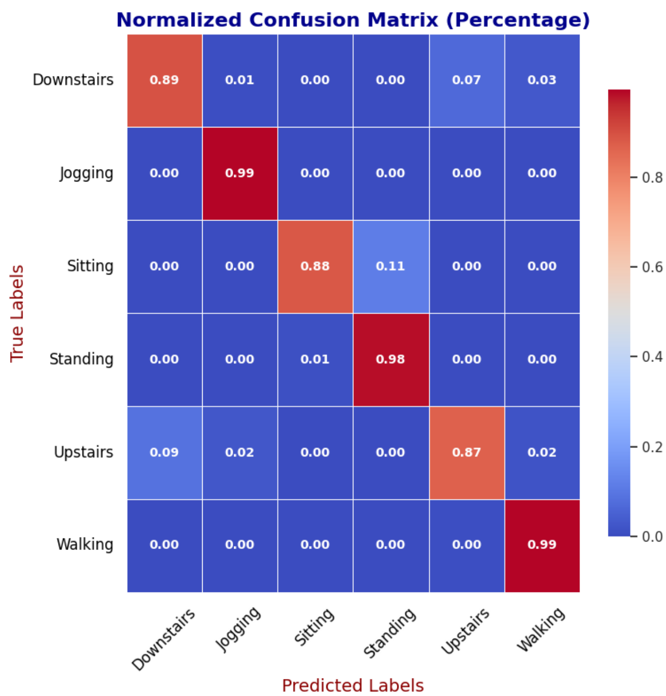


Fig. 3. Confusion matrix for Human Activity Recognition

capture both fine-grained and broader motion patterns.

1) Parameter Sharing: The feature maps are added directly by the addition layer, which promotes parameter sharing between the two pathways. This means that both pathways have an equal impact on the final result, and the model learns to combine information from both pathways more seamlessly.

2) Computational Efficiency: In comparison to concatenation, the addition operation is more computationally efficient and requires fewer parameters. This can be useful, particularly in situations where computational resources are limited for example Edge devices. Concatenation increases the dimensionality of feature maps, making them more computationally demanding.

3) Gradient Flow: The addition operation ensures stable training dynamics by promoting smoother gradient flow. The streamlined addition structure rules out the use of complex gradient pathways, which can stymie convergence, especially in deeper architectures. Conversely, concatenation results in a convoluted gradient flow, which exacerbates convergence issues and may impair the model's ability to learn effectively from training data.

C. Segmentation & Training

DNNs require fixed input sizes so the input signal must be segmented. The segmentation window's size is also significant. Empirical evidence suggests that a segment size of 256 with a step size of 16 yields optimal results for our tasks. After the creation of segments of the entire dataset, we divided the data into a 70% training set and a 30% testing set.

IV. Results

To evaluate our model, we use accuracy, precision, recall, and F1-score. Due to class imbalance, accuracy alone is insufficient, so supplementary metrics provide a more comprehensive performance assessment.

A. Human Activity Recognition

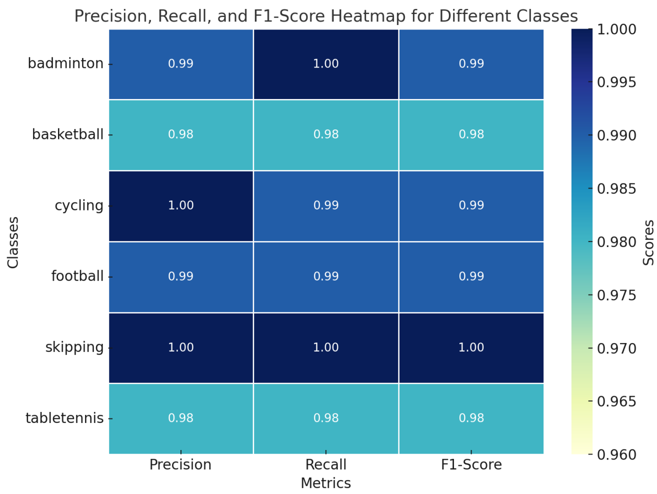
Spak-Net represents a significant improvement over its predecessor, Mukhtasir-Khail-Net. Spak-Net achieves a notable accuracy boost of 2.01% on the HAR WISDM-11 dataset. Spak-Net specifically attained an accuracy of 96.25%, surpassing Mukhtasir-Khail-Net's 94.24%. This advancement is further substantiated by Spak-Net's superior F1-score, precision, and recall metrics. Spak-Net's recall score for the 'Upstairs' class improved by 10%, from 77% to 87%. This substantial improvement demonstrates Spak-Net's capability to better capture complex activity patterns. Additionally, Spak-Net exhibited enhanced performance in other classes, such as 'Setting' and 'Walking', as illustrated in Figure 6. For a comprehensive understanding of Spak-Net's performance, we provide a detailed class-wise comparison of the performance report with Mukhtasir-Khail-Net in Figure 6. The confusion matrix in Figure 3 offers valuable insights into Spak-Net's Human Activity Recognition (HAR) capabilities. Spak-Net highlights its strengths and areas for further refinement.

B. Sports Activity Recognition

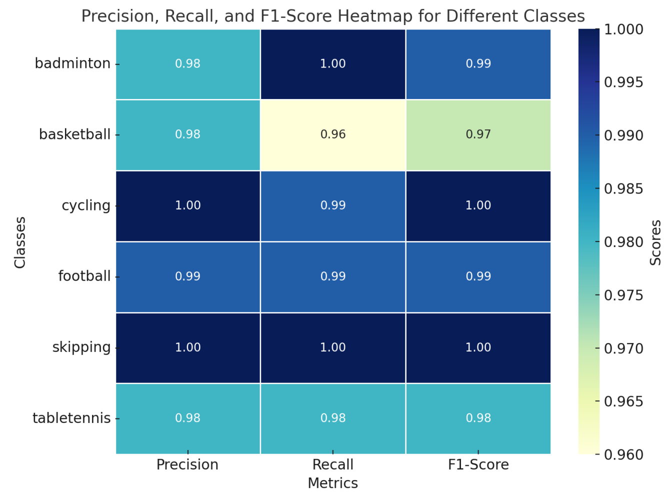
SpakNet demonstrates superior classification accuracy compared to Mukhtasir-Khail-Net. Spak-Net achieved an impressive 98.93% accuracy, outperforming its counterpart in clas-

This is the preprint version. Cite published paper: https://ieeexplore.ieee.org/document/11034469

This is the preprint version. Copyright belongs to IEEE. Kindly cite the published version: https://ieeexplore.ieee.org/document/11034469



(a) Spak-Net: Performance Report



(b) Mukhtasir-Khail-Net: Performance Report [3]

Fig. 4. Comparison of F1-Score, Recall and Precision for each class for Spak-Net and Mukhtasir-Khail-Net for Sports Activity Recognition

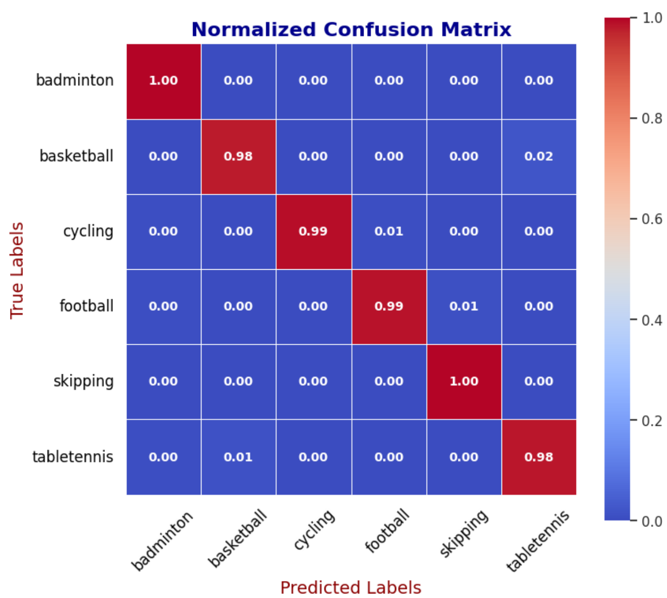


Fig. 5. Confusion matrix for Sports Activity Recognition

TABLE I
PERFORMANCE COMPARISON WITH EXISTING DEEP MODELS ON WISDM 11

Type & Reference	Accuracy (%)
[10] CNN	95.26
[17] LSTM-CNN	95.85
[9] CNN	94.65
[21] CNN	93.32
[22] CNN with an attention mechanism	96.40
[8] BiGRU-CNN	97.20
[23] CNN-BiGRU	98.81
[6] CNN (EdgeHARNet)	94.03
[5] CNN (Khail-Net)	89.60
[18] BiGRU	95.62
[3]Mukhtasir-Khail	94.24
Spak-Net	96.25

sifying sports activities, particularly 'Basketball' and 'Badminton'. A detailed comparison of the two models' performance is presented in Figure 1 4. The figure comprises two sub-diagrams: (a) Spak-Net's performance report and (b) Mukhtasir-Khail-Net's performance report. It is to be noted that Spak-Net excels in 'Basketball' classification, achieving a 98% F1-score and 98% recall rate, surpassing Mukhtasir-Khail-Net's 97% F1-score and 96% recall rate. Our model's simplicity is a significant advantage, boasting 1.6 times less complexity than Mukhtasir-Khail-Net. To provide additional insights, we have included the confusion matrix in Figure 5, which illustrates class-wise accuracy.

V. CONCLUSION

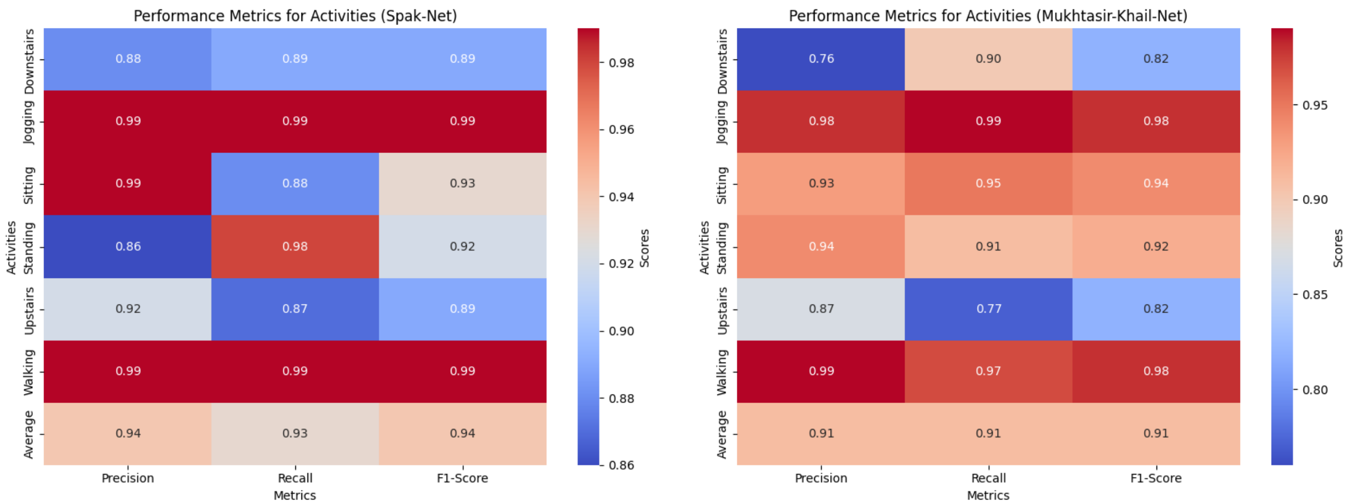
This research introduces a novel neural network architecture called Spak-Net. The model is an ultra-lean, convolutional-based design with only 398 trainable parameters. Spak-Net achieves superior performance in Human Activity and Sports Activity Recognition tasks. It surpasses its predecessor Mukhtasir-Khail-Net in multiple evaluation metrics. These metrics include accuracy, F1-score, precision, and recall. The model also reduces the size by a factor of 1.6 compared to Mukhtasir-Khail-Net. This innovative architecture combines computational efficiency with practical applicability. Spak-Net is particularly well-suited for real-time environments. It is ideal for edge computing applications where resources are limited and real-time computing is required. Spak-Net's performance may be affected by sensor noise and placement variations, which could be addressed in future work. Future work could explore adaptive learning techniques to enhance robustness in real-world conditions.

REFERENCES

[1] S. Muhammad, K. Hamza, H. A. Imran, and S. Wazir, "Enhanced human activity recognition using inertial sensor data from smart wearables: A neural network approach with residual connections," in *2024 International Conference on Engineering Computing Technologies (ICECT)*, 2024, pp. 1–6.

This is the preprint version. Cite published paper: https://ieeexplore.ieee.org/document/11034469

This is the preprint version. Copyright belongs to IEEE. Kindly cite the published version: https://ieeexplore.ieee.org/document/11034469



(a) Spak-Net: Performance Report

(b) Mukhtasir-Khail-Net: Performance Report [3]

Fig. 6. Comparison of F1-Score, Recall, and Precision for each class for Spak-Net and Khail-Net for Human Activity Recognition

[2] A. Jalal, M. A. K. Quaid, and A. S. Hasan, "Wearable sensor-based human behavior understanding and recognition in daily life for smart environments," in *2018 International Conference on Frontiers of Information Technology (FIT)*. IEEE, 2018, pp. 105–110.

[3] H. A. Imran, S. Muhammad, S. Wazir, A. A. Ikram, and O. Arshad, "Mukhtasir-khail-net: An ultra-efficient convolutional neural network for sports activity recognition with wearable inertial sensors," in *2024 4th International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, 2024, pp. 1–5.

[4] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner, and T. T. Pulickal, "Design considerations for the wisdm smart phone-based sensor mining architecture," in *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*, 2011, pp. 25–33.

[5] H. A. Imran, "Khail-net: A shallow convolutional neural network for recognizing sports activities using wearable inertial sensors," *IEEE Sensors Letters*, vol. 6, no. 9, pp. 1–4, 2022.

[6] H. A. Imran, A. A. Ikram, S. Wazir, and K. Hamza, "Edgeharnet: An edge-friendly shallow convolutional neural network for recognizing human activities using embedded inertial sensors of smart-wearables," in *2023 International Conference on Communication, Computing and Digital Systems (C-CODE)*, 2023, pp. 1–6.

[7] H. A. Imran, K. Hamza, and Z. Mehmood, "Harresnext: An efficient resnext inspired network for human activity recognition with inertial sensors," in *2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, 2022, pp. 1–4.

[8] H. A. Imran, "Ultanet: A antithesis neural network for recognizing human activity using inertial sensors signals," *IEEE Sensors Letters*, 2022.

[9] K. Mehmood, H. A. Imran, and U. Latif, "Hardensenet: A 1d densenet inspired convolutional neural network for human activity recognition with inertial sensors," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*. IEEE, 2020, pp. 1–6.

[10] H. A. Imran and U. Latif, "Hharnet: Taking inspiration from inception and dense networks for human activity recognition using inertial sensors," in *2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*. IEEE, 2020, pp. 24–27.

[11] A. M. Sabatini, G. Ligorio, A. Mannini, V. Genovese, and L. Pinna, "Prior-to-and post-impact fall detection using inertial and barometric altimeter measurements," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 24, no. 7, pp. 774–783, 2015.

[12] Q. Riaz, M. Z. U. H. Hashmi, M. A. Hashmi, M. Shahzad, H. Errami, and A. Weber, "Move your body: Age estimation based on chest movement during normal walk," *IEEE Access*, vol. 7, pp. 28 510–28 524, 2019.

[13] M. A. Hashmi, Q. Riaz, M. Zeeshan, M. Shahzad, and M. M. Fraz, "Motion reveal emotions: Identifying emotions from human walk using chest mounted smartphone," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 511–13 522, 2020.

[14] N. S. B. Suhaimi, J. Mountstephens, and J. Teo, "Emotional state classification with distributed random forest, gradient boosting machine and naïve bayes in virtual reality using wearable electroencephalography and inertial sensing," in *2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. IEEE, 2020, pp. 12–17.

[15] R. J. Kamil, D. Bakar, M. Ehrenburg, E. X. Wei, A. Pletnikova, G. Xiao, E. S. Oh, M. Mancini, and Y. Agrawal, "Detection of wandering behaviors using a body-worn inertial sensor in patients with cognitive impairment: A feasibility study," *Frontiers in Neurology*, vol. 12, p. 148, 2021.

[16] J. Qi, P. Yang, M. Hanneghan, S. Tang, and B. Zhou, "A hybrid hierarchical framework for gym physical activity recognition and measurement using wearable sensors," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1384–1393, 2018.

[17] K. Xia, J. Huang, and H. Wang, "Lstm-cnn architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56 855–56 866, 2020.

[18] K. Hamza, Q. Riaz, H. A. Imran, M. Hussain, and B. Krüger, "Generisch-net: A generic deep model for analyzing human motion with wearable sensors in the internet of health things," *Sensors*, vol. 24, no. 19, p. 6167, 2024.

[19] H. A. Imran, S. Wazir, A. J. Ikram, A. A. Ikram, H. Ullah, and M. Ehsan, "Hpc as a service: A naïve model," in *2019 8th International Conference on Information and Communication Technologies (ICICT)*, 2019, pp. 174–179.

[20] S. Wazir, A. A. Ikram, H. A. Imran, H. Ullah, A. J. Ikram, and M. Ehsan, "Performance comparison of mpich and mpi4py on raspberry pi-3b beowulf cluster," *arXiv preprint arXiv:1911.03709*, 2019.

[21] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Applied Soft Computing*, vol. 62, pp. 915–922, 2018.

[22] H. Zhang, Z. Xiao, J. Wang, F. Li, and E. Szczerbicki, "A novel iot-perceptive human activity recognition (har) approach using multihead convolutional attention," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1072–1080, 2019.

[23] H. A. Imran, Q. Riaz, M. Hussain, H. Tahir, and R. Arshad, "Smart-wearable sensors and cnn-bigru model: A powerful combination for human activity recognition," *IEEE Sensors Journal*, 2023.

This is the preprint version. Cite published paper: https://ieeexplore.ieee.org/document/11034469