Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 3, 2025

DEEP FUSION: A DEEP LEARNING FRAMEWORK FOR THE FUSION OF HETEROGENEOUS SENSORY DATA

Hamza Zahid[*1], Muhammad Arif[2], Hadia Hafeez[3], Muhammad Bilal Qureshi[4]

[*1,2,3,4]*Department of Computer Science & IT, Superior University, 10 KM Lahore- Sargodha Rd, Sargodha, Punjab 40100, Pakistan.*

[*1]hamzashaikhcan@gmail.com, [2]md.arif@superior.edu.pk, [3]hadiamehar09@gmail.com, [4]bilalshah1728@gmail.com

DOI: https://doi.org/10.5281/zenodo.15010209

Keywords
Deep learning, Sensor fusion, Human activity recognition, Multi-sensor data integration.

Article History
Received on 05 February 2025
Accepted on 05 March 2025
Published on 12 March 2025

Corresponding Author: *

Abstract

*Deep learning has revolutionized the integration of heterogeneous sensory data, facilitating the extraction of valuable insights from multiple sources. This study introduced **DeepFusion**, a deep learning-based framework aimed at enhancing multi-sensor data fusion by analyzing cross-sensor correlations and adaptively assigning weights according to measurement quality. The primary objectives were to develop an efficient sensor fusion model, evaluate its effectiveness in human activity recognition using real-world datasets, and compare its performance against existing approaches. While multi-sensor fusion improves accuracy, resilience to noise, and feature diversity, it also poses challenges such as increased computational demands and the need for data synchronization. To assess DeepFusion, two testbeds were constructed utilizing commercially available sensors, including smartphones, smartwatches, Shimmer sensors, WiFi, and acoustic sensors. Experimental evaluations revealed that DeepFusion outperformed leading methods in human activity recognition. The dataset encompassed various human activities in a device-free setting, such as typing, writing, and walking. Despite classification challenges observed in the confusion matrix, DeepFusion achieved the highest accuracy (0.908) on the CSI dataset, exceeding the performance of DeepSense (0.860), SR+WC (0.865), SR+Avg (0.833) and SVM (0.520), demonstrating its superiority in multi-sensor data integration.*

INTRODUCTION

The development of intelligent and user-friendly internet of things (IoT) system, based on interconnected computing and sensing devices, has garnered considerable attention in recent years. These systems paved the way for a new generation of applications capable of handling complex sensing and recognition tasks, enabling innovative interactions between individuals and their physical surroundings. The same object is typically being monitored by several separate sensors in many of these applications (Dhal et al., 2022; Dixit & Jangid, 2024; Weiberg & Finne, 2022). Wireless devices like laptops and iPads that are placed in the monitored subject's living quarters as well as gadgets, watches, smartphones and smart glasses—can yield valuable data about the subject's activities. Every one of these gadgets offered a different "view" of the object of observation and can be thought of as an information source. It seems sense that by combining the complimentary data from several sensors, it could

increase the accuracy of activity detection (Smith and Johnson, 2020). However, there are a number of issues that need to be resolved before it can fully utilized multi-sensor data. Initially, heterogeneous data may be provided by various sensors. On the one hand, many sensory data modalities (such as visible light, WiFi signal, acceleration readings and ultrasound) may be simultaneously collected in order to identify the same behaviours. However, different sensors may collect data in different ways (e.g., through varied transmission rates, signal strengths, or sampling rates), which will further heterogeneously integrate information gleaned from various devices. Second, the quantity of information that different sensors can capture varies depending on a number of factors, including hardware quality, placement, and ambient noise (Sachin & Jagdish, 2024; Alramadeen, 2022).

A perfect data fusion method should be able to distinguish between different sensors' varying levels of data quality and depend on the more illuminating ones. Third, there's a chance that the data gathered from several sensors will be associated with one another; for this reason, the data fusion model needs to record and account for this cross-sensor correlation (Jagdish, 2024 ; Qi et al., 2020). Address the aforementioned challenges, it was proposed to utilize deep learning techniques, which have demonstrated success in handling large, noisy and heterogeneous datasets. In this research, developed a deep learning system called DeepFusion to integrate disparate sensory data. Used a CNN-based Sensor Representation module in the model to uniformly reduce the dimensionality of diverse inputs while maintaining the distinctive qualities of each sensor view (Chen et al., 2019). A weighted-combination module by weighing the combination of multi-sensor features and assessing the value of evidence each sensor contributes. In order to extract and include cross sensor correlation features in research model, additionally build a Cross-Sensor module. By leveraging the multi-sensor structure, the suggested model can fully utilized the data gathered from various sensors with varying quality levels and can also identify distinct patterns in the data from various sensor perspectives (Bian et al., 2022).

1.1. Hybrid Fusion Networks

Hybrid fusion networks represent an effective approach for fusing heterogeneous sensory input by overcoming the weaknesses of conventional data fusion techniques. Such networks combine the best features of several architectures to efficiently integrate and fuse the input from different sources. The Attention-Based Two-Branch Hybrid Fusion Network, for instance, improves medical picture segmentation by integrating fine and coarse information across many scales through both CNNs and Transformers1. This dual-branch technique simultaneously extracts local and global information to improve the accuracy and robustness of the segmentation process (Chen et al., 2021).

A hybrid fusion network has been designed for the medical picture segmentation of detecting brain tumours. In this work, the multi-modal encoder-decoder architecture was used, which incorporates information from many modalities such as MRI and CT images. Adding a multi-modal hybrid fusion module enables the network to extract better distinctive features from each modality, thus reducing overall framework complexity while making it increase the accuracy of segmentation. Hierarchical and hybrid fusion networks have also been proposed as effective approaches to model motion dynamics and address non-homogeneous modalities. These networks enhanced performance in tasks such as autonomous driving and human activity recognition, where the two-stream variations are extended to three and six streams in order to be able to facilitate more complex cross-modal learning (Vasudeva Rao and Lingappa, 2019).

1.2. Networks, Graph Neural Networks (GNNs)

A specific type of neural network, called Graph Neural Networks (GNNs), is designed to process data organized as graphs, where nodes represent entities and edges represent the interactions between these things. GNNs are designed to handle non-Euclidean data structures, unlike traditional neural networks, which work on Euclidean data. This flexibility makes GNNs applicable in a wide range of applications. The fundamental mechanism of GNNs involves message passing in which nodes update their representation iteratively after having a discussion with neighbors over information exchange. This

mechanism allows the network to capture highly intricately related and interacting information within the data by means of aggregation as well as propagation of information throughout the graph (Liu and Zhou, 2022).

GNNs apply to the utilization of molecular biology, natural language processing, recommendation systems and studies of social networks and others. So many books introduce GNN in a way as comprehensive as describing designs, major ideas, and realistic applications. And Towards AI studies the theory on GNN for designing complicated relationships. Generally speaking, the series above describe very broadly how data is handled based on the foundations of GNN (Zhou et al., 2020).

1.3. Deep Belief Networks (DBNs)

One variety of deep neural networks called the deep belief networks that consists of many layers of hidden units, known as latent variables and the networks are excellent in learning hierarchical data representation, which is why they are very suitable for applications like natural language processing, audio recognition and image recognition (Hinton et al., 2006). Usually, restricted Boltzmann machines are stacked into layers to construct DBNs with each layer trying to extract even more extract features from the input data. DBNs trained into 2 strategies: Pre-training, in which each layer is taught individuals and unsupervised to determine probability distribution of data and Fine-tuning, in which network as a whole is trained supervised to improve performance on certain tasks (Mohamed et al., 2011). DBNs can captured deep data representation and described complex data distribution, they are widely employed. They have shown very good results in voice recognition, where they model temporal correlation in audio signals to improve speech to text accuracy and image classification where they automatically learn features from raw pixel data. DBNs also used for natural language processing such as machine translation and sentiments analysis because they are efficient at capturing complex words phrase association (Bengio, 2009).

This study presented a DeepFusion architecture specially designed for a sample sensing task: human activity recognition (HAR). HAR is a critical component in various IoT applications, including smart homes, healthcare systems and fitness tracking. Although the primary emphasis was on the HAR application, the framework versatile design holds promise for broader IoT applications involving classification or identification tasks. Comprehensive real-world experiments were conducted to evaluate the DeepFusion framework across both device-free and wearable human activity recognition scenarios. The results demonstrated the effectiveness of the proposed model, showcasing notable improvements over existing state-of-the-art algorithms. The key assistances of the study can be briefed as follows:

- Listed both the advantages and disadvantages of fusing diverse multi-sensor data.
- DeepFusion is what was proposed to develop informative representations of heterogeneous sensory input. To improve the sensing performance, DeepFusion may compute the crosssensor correlations and pool data from different sensors by assigning appropriate weights to every sensor's measurement based on the quality of the measurement.
- Developed two testbeds using available COTS equipment, including wearable technologies from smartphones, smartwatches and Shimmer sensors, and wireless sensing devices from WiFi and acoustic sensors. Real-world human activity data was collected and empirical evaluation showed that the proposed DeepFusion model outperformed the stateof-the-art methods in human activity recognition when applied to this dataset.

2. LITERATURE REVIEW

Guo et al. (2019) proposed the iFusion framework to address the challenge of real-time integration of disparate data sources for deep learning. It employed advanced data fusion algorithms to process different kinds of data such as text, images and sensor data efficiently. Applications that require fast decisions, such as autonomous driving, healthcare and smart cities, depend on the real-time efficiency of iFusion. Due to iFusion, which integrates pre-processing, feature extraction and training in a data fusion model, deep learning models can learn from integrated data efficiently. This methodology improved fast and accurate decision-making across multiple domains and represented a significant advancement in data fusion for deep learning.

2.1. Deep Fusion and IoT-Based wearables

Bahador et al. (2021) focused on deep learning for multimodal data fusion by detecting food intake using wearable sensors. They have developed an efficient technique to deal with high-dimensional multisensory data by transforming time-series data into 2D space for the better classification of eating episodes. This has helped in improving accuracy and scalability in the integration of sensor data, allowing for personalized health monitoring. This work contributes to further advancements in wearable technology and real-time data processing in health informatics.

Machine learning (ML) has gained recent success in a wide range of fields, especially in the medical and bioinformatics fields. Researchers have conducted extensive studies on the application of deep learning (DL) techniques to solve the problems in these fields, which is important because ML is very precise. Deep learning's ability to process and interpret vast, complex and diverse datasets in realtime is especially valuable for bioinformatics and medical applications within Internet of Things (IoT) systems. This capability provided insights that can improve healthcare outcomes and enhance operational efficiency across the industry. In IoT-based bioinformatics and medical informatics, DL has diverse applications, including image analysis, wearable device monitoring, clinical decisionmaking, diagnostics, therapy recommendations and drug discovery (Amiri et al., 2024).

Dargazany et al. (2018) introduced Wearable Deep Learning (WearableDL), a unified conceptual framework that integrates wearable technologies (WT), the Internet of Things (IoT) and deep learning (DL). This architecture addressed key challenges by: (1) ensuring scalability to manage large datasets; (2) enabling autonomous feature engineering without manual feature extraction or handcrafted features; and (3) achieving high accuracy and precision when learning from both raw labeled and unlabeled data (supervised and unsupervised learning). In an effort to grasp the current state of affairs, the authors conducted an extensive review of over 100 recently published studies focused on developing DL algorithms for wearable and user-centered technologies. Their findings reinforced and refined the proposed bioinspired WearableDL architecture. The study also provided valuable insights and practical recommendations on the application of WearableDL in big data analytics.

2.2 Framework for the Fusion of Heterogeneous Sensory Data

Ignatious et al. (2023) research introduced an adaptive selective sensor fusion paradigm for enhancing the robustness of autonomous vehicles (AVs) under challenging driving scenarios. The proposed framework adapts sensor fusion at runtime to suit the driving environment through switching between early, late, and hybrid fusion approaches. It is composed of components for advanced object detection/classification, feature selection and the processing of large amounts of different types of data. Comparing the results of experiments with existing fusion models, such as KITTI and nuScenes, reveals higher accuracy and efficiency.

Liu et al. (2017) also pointed out the increasing interest in heterogeneous sensor data fusion while considering it as a complex and challenging area. Main challenges dealing with missing values in datasets and creating shared representations for multimodal data in order to improve inference and prediction accuracy. To address these challenges, a Deep Multimodal Encoder (DME) was proposeda deep learning-based system designed for tasks such as novel modality prediction, sensor data compression and imputing missing values in multimodal environments. DME successfully captures intermodal correlations at deeper network layers, while intramodal correlations are focused in the initial layers. This makes it superior to traditional methods, which mainly emphasize intramodal relationships. DME was proven to be highly effective in experiments using real-world data from a 40-node agricultural sensor network with three modalities, such that the root mean square error for missing data imputation was a mere 20% of that obtained by conventional approaches like K-nearest neighbors and sparse principal component analysis.

2.3 Deep Multimodal Fusion of Data with Heterogeneous Dimensionality

Li et al. introduced a new architecture of deep learning for combining multimodal data of different dimensions in 2022. It extracted features from

multiple modalities and projected the features into a shared feature subspace to enhance the effectiveness of data fusion. The method overcomes challenges with heterogeneous dimensionality, leading to accurate and reliable localization. The authors evaluated their strategy using experiments and confirmed that it outperformed state-of-theart methods in accuracy and robustness. This approach emphasizes the value of integrating many data modalities and is especially helpful for robotics, autonomous vehicles and augmented reality applications.

Morano et al. (2024) studied diagnosis and treatment of many diseases have significantly improved as a result of the use of multimodal imaging. Some works have shown the advantages of multimodal fusion for automatic segmentation and classification utilizing deep learning-based techniques, which is comparable to clinical practice. Nevertheless, the fusion procedures used by classification systems are incompatible with localization tasks and existing segmentation approaches are restricted to the fusion of modalities with the same dimensions (e.g., 3D + 3D, 2D + 2D), which is not always feasible. The suggested framework projects the features into the common feature subspace by extracting them from the various modalities. The framework was tested on two tasks: segmenting retinal blood vessels (RBV) in multimodal retinal imaging and segmenting geographic atrophy (GA). According to their study findings, the suggested approach performs up to 3.10% and 4.64% Dice better on GA and RBV segmentation, respectively, than the most advanced monomodal approaches.

2.4 Optimization of Data Fusion in Industrial Environments

Deng et al. (2019) presented hybrid framework integrating CNNs, RNNs and DNNs for data fusion and anomaly detection in industrial settings. The system improves model accuracy and resilience by using CNNs for feature extraction from geographical data, RNNs for temporal data, and DNNs for feature integration. It effectively handled high-dimensional data processing from multiple sources, including cameras and sensors. Experiments conducted on real-world industrial datasets show notable gains in anomaly detection accuracy over baseline techniques. All things considered, this study demonstrated how deep learning methods may be used to optimized data fusion in industrial settings, resulting in more efficient control and monitoring systems.

Sultani et al. (2014) explained that Wireless Sensor Networks (WSN) are still expanding at an astonishing rate. WSNs are being used in a variety of fields, including environmental, medical and transportation as well as military and transportation applications. Numerous moved may be present in some of these applications, which presents serious problems for data transfer, network longevity and overall reliability. Data fusion techniques are becoming more and more popular in WSNs to increase the accuracy of reported data and aid in event prediction. They are employed to raise the information's level of dependability. Although data correctness is addressed, the inefficiencies brought forth by very big nodes and excessive data redundancy remain unaddressed. While data aggregation is a straightforward method of optimizing data flow, it is not a comprehensive solution. The huge WSN size reduces the performance of the WSN and may even completely impair its operation due to congestion and increased traffic load in the network.

2.5 Data Fusion for Human Activity Recognition Using Deep Learning

Vidya and Sasikumar (2022) concluded, with wearable sensors, HAR has many applications in the field of smart homes, surveillance, fitness and healthcare. In spite of the vast computational research on Human Activity Recognition (HAR), many issues are still open for solving multi-sensor-based activity recognition. Such issues involve handling complex time series data, extracting the meaningful feature vectors from multimodal data and dimensionality reduction, all of which have to be explored more. Four machine learning (ML) classifier models, namely Support Vector Machine (SVM), KNearest Neighbor (KNN), Ensemble Classifier (EC) and Decision Tree (DT) were trained to classify various human activities. This was done by using entropy features derived from Empirical Mode Decomposition (EMD) and discriminative statistical information obtained from Discrete Wavelet Transform (DWT). The proposed method was

validated using a publicly accessible UCI dataset. The experimental results, analyzed via confusion matrix and parallel coordinate plot (PCP), showed that the ML-based Human Activity Recognition (HAR) framework reached up to 99.63% accuracy in classification.

2.6 Deep Learning Techniques for Sensor Data Fusion in IoT

Rajawat et al. (2021) investigated several deep learning techniques for sensor data fusion in Internet of Things applications and highlighted the advantages, disadvantages, and future prospects of each technique. The authors categorize these techniques-which are autoencoders, CNNs and RNNs-and provide a comprehensive overview of their Internet of Things applications. Aiming at improving the accuracy, resilience, and scalability of IoT systems, the survey points out the importance of sensor data fusion and tackles issues such as data heterogeneity, real-time processing, and energy efficiency. It thus highlights important topics for additional research and provides insightful information on the state of the field. Krishnamurthi et al. covered a number of real-world issues, including issues related to smart cities, healthcare, building management, transportation, and environmental monitoring, according to 2020.

The dataset from Duvall et al. (2016) contained Ozone (O3) and Nitrogen Dioxide (NO2) concentrations measured in Houston, Texas, from 4–27 September 2013 using Cairclip sensors and Federal Reference Monitors (FRM). O3 values were derived by subtracting NO2 values from the CairclipO3/NO2 sensor. O3 data showed good agreement with reference instruments (r=0.82), but NO2 data exhibited low agreement (r=0.08).

2.7 Food Intake Episodes Detection Using Wearable Sensors

Bahador et al. work of the year 2021 reflects on the use of deep learning-based multimodal data fusion algorithms and wearable sensors for detection of food intake events. The time-series data was transformed to 2D space in order to allow proper categorization and, in fact, the authors were able to manage the computationally complex high-dimensional data coming from different sources. This methodology leverages statistical dependencies and correlations across disparate sensory inputs. It yields more specific insights into dynamics in human behaviour. The context of the presented study showed method applicability within many contexts along with promising personal eating habit tracking. The major findings of this study underpinned the significance of multimodal data fusion in terms of wearable technologies for health monitoring and the possible resolution of existing problems related to heterogeneous data integration and real time data processing.

Bedri et al. (2017) solved the problem of food journaling that is often used but often undermined by self-bias and recall errors leading to poor adherence in the users. The wearable device EarBit was designed to detect mealtimes. In the study, the efficiency and usability of inertial, optical and acoustic sensing modalities have been studied in detail, with a greater focus on inertial sensing. The results showed that EarBit attained an F1-score of 90.9% and recognition accuracy of 90.1% in highly controlled laboratory settings. In a natural, real-world unconstrained setting, EarBit correctly detected chewing instances at a 93% level with an F1-score of 80.1% while detecting nearly all episodes of eating.

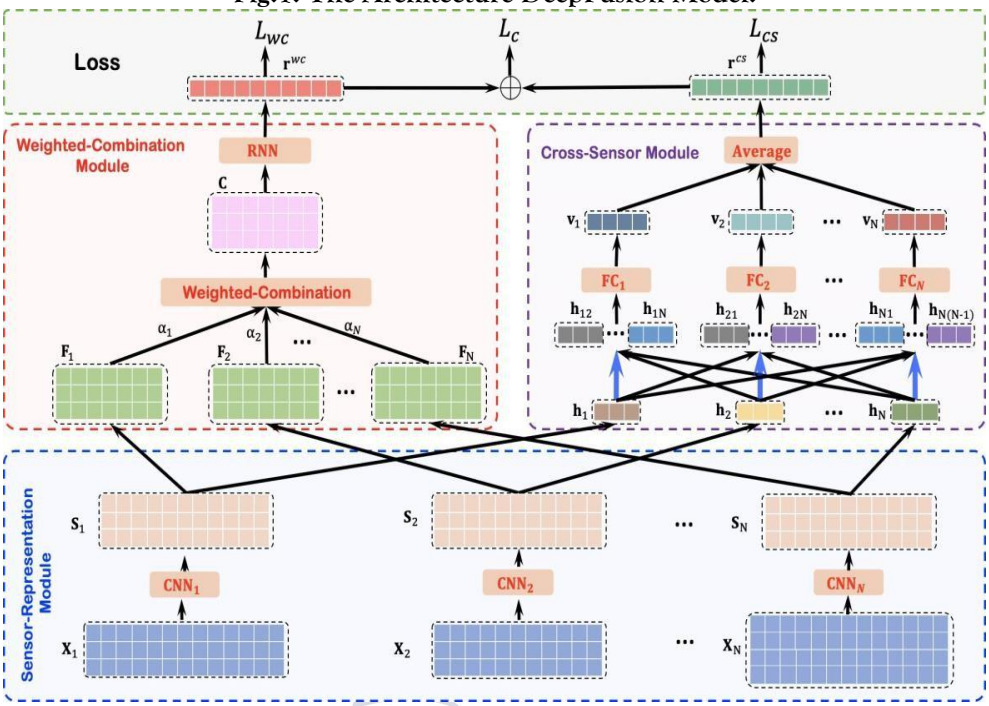
Critical for treating medical conditions like obesity and eating disorders is Fontana et al. (2021) research on understanding people's ingested behavior. Giving the energy and minerals needed for human existence also requires food intake. Historically, ingestive behavior has been evaluated through self-monitoring of food intake; however, this approach is often imprecise, laborious and prone to misreporting. An encouraging alternative are wearable sensors, which monitor physiological responses related to several stages of food consumption: hand-to-mouth movements, biting, chewing and swallowing-to provide objective measurements. Sensor data is processed using advanced signal processing and pattern recognition algorithms to automatically identify and document each eating experience.

3. METHODOLOGY

This section introduced Deep Fusion, a unified deep framework for human activity recognition using multi-sensor data, designed to handle heterogeneous inputs effectively. The architecture of the proposed

model, depicted in Figure 1, consists of three core modules: the sensor representation (SR) Module, the Weighted Combination (WC) Modules and the Cross-Srnsor (CS) Module. Each module is explained in detail in the subsequents. Throughout the framework, bold uppercase letters such as weight matrix W represent matrices and tensors. Bold lowercase letters, such as bias vector b, demoted vector and standard characters such as hyper parameter *a* indicate scalar values.

Fig.1. The Architecture DeepFusion Model.



3.1 Sensor-Representation Module

The multi-sensor data collected is represented as a set of heterogeneous continuous time series, each comprising multiple non-uniformly sampled signals. A CNN-based module is introduced to directly learn sensor representations from the raw heterogeneous data in order to retain the distinct features of each sensor while standardizing the input dimensions across sensors. Convolutional Neural Network (CNN) blocks are particularly well-suited for this activity recognition framework due to their efficiency and effectiveness. These blocks convert raw data into low-dimensional representations while preserving their original size, thereby enhancing the capability to express features effectively.

In a Convolutional Neural Network (CNN), convolutional layers with flexible filters and strides play a pivotal role in reshaping input data. Additionally, pooling layers may be employed for downsampling the data. When constructing stacked CNN blocks for different sensors in this model, several parameters must be defined. Assuming the input data comprises N sensing nodes, denoted as $\{X_1, ..., X_i, ..., X_n\}$, the sensor representation for the *i*-th sensing node (X_i) is obtained through a series of stacked CNN blocks, represented as CN N_i.

$Si = CNNi (Xi ; \theta i)$

The function CNNi was the stacked CNN blocks that are tailored for each input matrix Xi and θi denoted the parameters to be learned for these specific stacked CNN blocks. In this model, it was also utilized the ReLU activation function, batch normalization and dropout techniques within the CNN blocks. The sensor representations obtained from all sensing nodes, denoted as {S1,. . ., Si,. . ., SN}, are exploited to characterize the heterogeneous inputs within a high-level feature space of uniform size.

Next, to standardize feature extents for input into the remaining two modules, the learned hidden representations were reformatted by combining and flattening processes. For the input to the WC Module, the pooling operation was applied by way of follows:

Fi = Pooling (Si)

Above formula allowed to reduce the size of the input features to an appropriate level for calculating numerical weights, which are subsequently passed to RNN. In contrast, for the Cross-Sensor Module, utilized the flattening operation to create a cross-sensor input vector hi for the i-th sensor, as follows:

hi = Flatten(Si)

By adopting this approach, the model efficiently handles the varying dimensionalities of raw data from different sensors. This helped apply dedicated CNN blocks to each sensor, each with its unique set of parameters, in order to standardize heterogeneous inputs and this method enabled the model to learn meaningful representations from sensors with the hope that there would be preservation of the distinct characteristics of each sensor in addition to enhancing performance.

3.2. Weighted-Combination Module

Different sensors may provide different levels of information due to the type of sensing signal, hardware quality, distance and angle to the observed object, as well as environmental conditions and ambient noise. An effective activity recognition method needs to address these differences in data quality across sensors and prioritize the more informative ones to enhance overall performance. To resolve this issue, a Weighted-Combination module was introduced. This module determines the quality of information from each sensor, which is called the quality weight and integrates the multisensor data in a weighted manner.

The Weighted-Combination module is inspired by the attention mechanism (Firat et al., 2016), a weighted aggregation technique commonly used in applications such as machine translation, computer vision and disease prediction (Yuan et al., 2018). However, traditional attention mechanisms often assume that only a few views are relevant to the task, which results in most views receiving near-zero weights. This assumption does not stand for human activity recognition, which may involve providing multiple sensors' valuable insights collectively to enhance the performance.

To tackle this challenge, introduced a novel weight-assignment strategy aimed at maximizing the use of multi-sensor information. For the learned weight-combination input matrices {F1, · · · , Fi, · · · , FN}, first utilized a pooling layer followed by the flattening operation to derive their encoding vectors {u1, · · · , ui, · · · , uN} for each sensor. Specifically, the encoding vector for the i-th sensor, ui can be computed as follows:

ui = Flatten (Fi)

Subsequently, the quality weight for the i-th sensor, denoted as ei, can be calculated using the following formula:

$e_i = w^{wc\,T} u_i + b^{wc}) / {}^{l}w^{wc}$ (eq1)

where w^{wc} and b^{wc} are the parameters that need to be learned and lw cl represents the length of the encoding vector ui. Based on Equation (1), it can then derived a normalized quality weight αi as follows:

$$\alpha_i = \frac{\tilde{\alpha}_i}{\sum_j \tilde{\alpha}_j}$$

Where $\tilde{\alpha}i$ represents the rescaled quality weight obtained using a sigmoid-based function.

$$\tilde{\alpha}I = \frac{a}{1+\exp\left(-\frac{ei}{b}\right)} + c \qquad \text{(eq 2)}$$

Where a, b and c are predefined hyper parameters. In Equation (2), the upper and lower bounds of the rescaled weights are a+c and c, respectively, while b controls the slope of the function near zero. By appropriately setting these hyper parameters, the variance of the normalized quality weights among all sensors can be minimized, allowing our model to utilize a greater number of sensors for activity recognition. Using the normalized quality weights of all sensors {α1,⋯ ,αi,⋯ ,αN}, the sensor combination matrix C can be calculated by weighted aggregation.

$C = \sum \alpha i \odot Fi$ (Eq 3)

⊙ denotes element-wise duplication. The sensor combination as a vector, use a 2-layer stacked Gated Recurrent Unit (GRU) to compute the output vector of the Weighted-Combination Module, denoted as per r^{wc} as follows:

$H_{1:L}=GRU(C_{1:L;\phi})$ (Eq 4)

$r^{wc}= \sum l\, H_l$ (Eq 5)

L represented the column length of the matrix, Hi denoted the output vector from GRU and ϕ encompasses all the GRU parameters. The output vector rwc is obtained by summing the output vector from GRU. This method enabled the model to effective utilized multi-sensor information by

capturing the variability across different sensors and emphasizing those that provide more significant data.

3.3. Cross-Sensor Module

The previous module successfully combines multi-sensor data but treats each sensor independently, neglecting the correlations between them. In human activity recognition, different sensors often provide related information that can enhance the overall understanding of the activity. By capturing and incorporating cross-sensor correlations into the deep learning framework, the model can identify more generalized patterns and improve its performance. This could lead to merely concatenating raw input features that might not provide robust or accurate results. In order to surmount this limitation, a Cross-Sensor module was introduced to add complementary information to the features obtained from the weighted combination module.

In this module, given the cross-sensor input vectors {h1,⋯,hi,⋯,hN}, the correlation vector for the ith sensor computed as follows:

$v_i = f (W^v i ^T [h\tilde{} _{i,1} \oplus h\tilde{} _{i,2} \oplus \cdots \oplus h\tilde{} _{i,N}] + b ^v i)$

$\oplus$ represents the concatenation operator, W^v i and b ^{v} i are the learnable parameters of FCi ^{cv} (a singlelayer fully connected neural network) and h ij denoted using element-wise variance (Mou et al., 2015).

$h\tilde{} _{i,j} = h_j - h_i$

It is essential to highlight that there are N−1 consistent associations, excluding self-correlation. By utilizing the correlation vectors {v1,...,vi,...,vN} the output vector of the Cross-Sensor Module, represented as r ^{cv}, can be obtained through an averaging operation.

$r^{cv} = 1 / N \sum Vi$

This allowed the model to capture the correlations between multi-sensor data in a low-dimensional space. Unlike the WC Module, whose features are integrated together with weighted combinations from different sensors, the Cross-Sensor Module makes use of an averaging operation. This is because each correlation vector already inherently contains the relationships between sensors and does not necessarily need them to be assigned varying importance factors.

3.4. Sample Population

Data was collected from publicly available IoT datasets in CSV (Comma Separated Values) format from platforms such as:

- Kaggle
- UCI Machine Learning Repository
- ThingSpeak
- Microsoft Azure Open Datasets
- AWS Public Datasets • Google Dataset Search
- CityPulse Dataset Collection
- OpenAIRE These datasets contained sensory information across various domains, such as smart cities, healthcare, environmental monitoring and human activity recognition, which was analyzed to evaluate the proposed deep learning framework.

3.5. Data Preparation

The data preparation phase was involved in cleaning and pre-processing the collected datasets. This step was included:

- Removing duplicates: Ensuring no repeated entries exist in the data.
- Handling missing values: Using imputation methods or removing entries with missing data.
- Data transformation: Changing raw data into a structured format appropriate for deep learning examination.

Additionally, each dataset was structured to create unique identifiers for sensors, timestamps and the associated environment. Feature extraction was performed to prepare data inputs for algorithms like Hybrid Fusion Networks, Graph Neural Networks (GNNs) and Deep Belief Networks (DBNs).

3.6. Framework Development

The framework was developed based on the combination of Hybrid Fusion Networks, GNNs and DBNs. The following steps was followed for framework implementation:

3.6.1. System Initialization:

Cryptographic keys were generated with the process of system initialization, which was a must setup safe data handling protocol. These keys made it possible for encrypted communication to occur while ensuring sensitive information during its transmission remains private. The framework also combines strong algorithms that build on data

authenticity and integrity while protecting illegal access and thefts of data. This phase also arranges in-depth validation procedures with the purpose of verifying the user and device identification in the system. Finally, these basic steps form the outline of a sound and reliable system for data management.

3.6.2. Sensor Data Management:
Hybrid fusion network were used to implement sensor data management which aggregated data from multiple sensors in an intelligent manner. To ensure that the most trustworthy and pertinent data was incorporated into system, the fusion process gave priority to the quality and kind of data. The environment was seen in its entirety and cohesively by combining sensor data from several sources.

3.6.3. Graph Neural Networks:
To capture cross sensor correlations appropriately, the complicated interaction between sensors were modelled using graph neural network (GNNs). GNNs might be examined and comprehend the relationship and dependencies between sensors by expressing them as nodes and edges inside a graph structure. Using this methods, the system was able to take into consideration both direct and indirect influences between sensor data, which increased forecast accuracy and robustness. The system capacity to handle linked data streams was improved by GNNs, leading to more thorough insights and effective decision-making.

3.6.4. Deep Belief Networks:
DBNs allowed for the extraction of hierarchical features from fused sensor data in such a way that deep insights into intricate patterns were gained. This was made possible due to their ability to learn low-level and high-level representations using many layers of obstruction that increase the significance of the data for tasks that came after. These extraction features improved the system capacity to make precise predictions and classifications, particularly in situations involving sizable and varied datasets. The most important elements are given priority. It promotes more effective and efficient decision making.

3.7. Data Analysis
3.7.1. Performance Evaluation:
The framework underwent rigorous testing since it was combining different sensory data and enhancing functionality in internet of things applications. The system managed to provide more dependable and comprehensive analysis through the use of data by many sensors for integration. These metrics included information on how well the system could anticipate and categorized events reduce false positives and negatives and strike a balance between precision and recall. The outcomes showed how much framework may improve the functionality of internet of things applications.

3.7.2. Scalability and Reliability Testing:
For testing system stability and reliability in dealing with large data sets of various sources, comprehensive stress tests were performed. This step aimed at determining the system's presentation under extreme conditions by generating large data volume and traffic loading. The key focus of this stage was to ensure the framework was robust enough to process data in real time without losing the accuracy and performance. Stress testing also helped identify potential bottlenecks ensuring that the system could scale without interruption while retaining its robustness and dependability in extensive IoT scenarios.

3.8. Reporting and Documentation
All findings from the framework's implementation and performance evaluation was compiled into a comprehensive report. That was included:
- Detailed explanations of the methods used.
- Performance metrics and usability assessment.
- A final framework documentation outlining design, implementation and user guidelines.

This research contributed to the growing wealth of knowledge in deep learning-based sensor fusion by offering informative information on how complex algorithms can be applied to fuse data from different sensor sources. The study became a driving force in furthering research into large-scale IoT systems as it demonstrated the efficiency of methods such as Graph Neural Networks and Hybrid Fusion Links. These results did not only suggest that IoT applications might perform better but also served as

a motivation for future works to develop more effective, reliable and flexible sensor fusion frameworks.

4. FINDINGS

The DeepFusion model was evaluated using real-world test beds for human activity recognition. This started with the state-of-the-art methods for human activity recognition used as baseline comparisons. Then experiments were conducted on human activity data gathered from two test beds: using commercial off-the-shelf wearable devices such as smartphones, smartwatches and Shimmer sensors, and also wireless sensing devices such as Wi-Fi and acoustic sensors.

4.1. Baselines

The SVM is a widely administered machine learning model, as evidenced in previous research by Zhou et al. (2017), who used it for human activity recognition tasks. Since a standard linear SVM is built for binary classification, a one-vs-all strategy was used in this study to solve the multi-class classification problem. For the experiments, data from the whole sensors was merged into single flattened article vector, which later was put to the SVM model.

DeepSense is one of the state-of-the-art deep learning models for the classification of multi-sensor data. Its architecture consists of three local CNN layers, three global CNN layers and two GRU layers (Ali et al., 2015). In these experiments, followed the settings used in the original paper. Specifically, each convolutional layer used 64 filters of size 3 × 3. Furthermore, dropout and batch normalization techniques are used to enhance the performance of the model.

Variants of DeepFusion: The proposed model of DeepFusion takes both the different contributions of a variety of sensors and the correlations into account. Its three main modules are: Sensor respresentation, Weighted combination module and Cross sensor module. To set baselines. 4.2. Experiments on Wearable Sensor Data In this section, assessed the performance of the proposed DeepFusion model using a real-world activity dataset gathered from various wearable sensors positioned on different parts of the body.

4.2.1. Experiment Setup

Three types of wearable devices were used, namely smartphones, smartwatches and Shimmer sensors. Triaxial accelerometers, gyroscopes and triaxial magnetometers were each attached to them. Data collected from six different volunteers, all males and females, who are wearing four different sensors on body regions: namely, two placed on the upper left arm; one on the left waist and one on right wrist and, finally, another on the right ankle. A total of 27 activities were conducted, as indicated in Table 1, where each participant was required to carry out each activity for one minute.

Table 1: Activities in Wearable Sensor Dataset.

No.	Activities	No.	Activities	No.	Activities
1	running	10	going downstairs and making a phone call	19	standing and washing hands
2	running in place	11	going upstairs	20	standing and wiping the blackboard
3	sitting and making a phone call	12	going upstairs and making a phone call	21	standing and wiping the table
4	sitting and keyboarding	13	standing and making a phone call	22	standing and writing
5	sitting and typing on the phone	14	standing and washing the dishes	23	standing and writing on the blackboard

6	sitting still	15	standing and keyboarding	24	walking backward
7	sitting and wiping the table	16	standing and typing on the phone	25	walking and making a phone call
8	sitting and writing	17	standing still	26	walking forward
9	going downstairs	18	standing and brushing the teeth	27	walking in place

4.2.2. Data Preprocessing

The data collected from each sensing device includes nine signals: three from the accelerometer axes, three from the gyroscope axes and three from the magnetometer axes. Despite variations in sampling rates and value ranges among the sensors, all data was downsampled to 25Hz and scaled to a range between 0.0 and 1.0 based on their magnitudes. The data was then segmented into non-overlapping 2-second windows, each containing 50 data points. Each segment was paired with its Fast Fourier Transform (FFT) to serve as input for the deep learning model. As a result, the final dimensions of each data segment from a single sensor are $9 \times 50 \times 2$. For the traditional model based on classification by Support Vector Machines (SVM), each signal provided by the accelerometer, gyroscope and magnetometer on every sensor was considered and for all of these 36 features, 432 were derived. The extracted features comprised of mean, standard deviation, MAD, median, maximum, minimum, energy, signal magnitude area and interquartile range of each axis for x, y and z axes. Additionally, magnitude of each signal was calculated together with the angles between each signal and its three axes, pairwise correlations among the axes, the energy of each signal and the signal magnitude area.

4.2.3. Model Settings

In the experiments involving wearable sensor data, constructed a sensor representation extractor consisting of six stacked CNN blocks for each sensing device. The convolutional layers utilized filters of sizes 3×5, 3×3, 3×3, 3×3, 1×3 and 1×3, with each convolutional layer containing 64 filters. The initial four CNN blocks operate without padding and max pooling is applied to down sample the data. For the Gated Recurrent Unit (GRU), the hidden state size is configured to 64. In the fully connected neural networks of the Cross-Sensor Module, the size of the shortened link vector vi also set to 64. ReLU is used as the activation function throughout. Dropout rates are set to 0.8 for the CNN and 0.7 for the RNN. The hyper parameters are configured as follows: a=9.0, b=0.01, c=10.0, β=0.1and γ=0.1. The model training process utilizes the ADAM optimization algorithm with a learning rate of 1e−4 and a batch size of 100. Performance is assessed based on accuracy.

4.2.4. Performance Validation

In this experiment, a leave-one-subject-out strategy was used for the evaluation dataset and the average accuracy score across all subjects was calculated as the performance metric. Table 2 presents the accuracy results for all methods applied to the wearable sensor data. The proposed DeepFusion model achieved the highest performance, while the traditional Support Vector Machines (SVM) approach performed the least effectively. This highlighted the superiority of deep learning models in human activity recognition (HAR) tasks. Among the three deep learning baseline models, SR+WC achieved the highest accuracy, due to its consideration of data quality from different sensors, unlike DeepSense and SR+Avg, which do not account for variations in data quality across sensors. The DeepFusion model effectively weights different sensors and captures their interrelationships, leading to the best performance.

Table 2. Performance on the Wearable Sensor Data.

Model	Accuracy
SVM	0.350
DeepSense	0.862
SR+Avg	0.835
SR+WC	0.870
DeepFusion	0.905

4.3. Experiments on Device-Free Human Activity Data

While device-based methods effectively monitor human activities, they come with significant limitations, such as the added burden and discomfort for users who must wear the devices. To overcome this challenge, considerable efforts have recently been directed toward developing devicefree human activity recognition techniques that utilize information from existing indoor wireless infrastructures, eliminating the need for individuals to carry dedicated devices. The underlying principle of these methods is that a person's activities can be inferred by analyzing the information conveyed through wireless signals transmitted between paired devices (e.g., smartphones, laptops, WiFi access points). Each sender-receiver pair provides a distinct "view" of the monitored subject.

4.3.1. Experiment Setups

The experiment involved analyzing seven distinct human activities, as detailed in Table 3. Data collection was conducted with eight participants, including both males and females. Each participant performed each activity for a duration of 51 seconds, with all activities repeated across two rounds. Two types of signals, WiFi and ultrasound, were recorded during the process.

WiFi signals were captured using a TP-Link AC3150 Wireless WiFi Gigabit Router (Archer C3150 V1), which transmitted packets to multiple receivers at a steady rate of 30 packets per second, simulating typical real-world wireless communication. Each receiver was equipped with an Intel Wireless Link 5300 NIC, operating on Ubuntu 11.04 LTS with a 2.2.36 kernel. The receivers utilized the Linux 802.11n CSI extraction toolkit, which generated Channel State Information (CSI) matrices for 30 sub-carriers across both the 2.4 GHz and 5 GHz frequency bands.

For ultrasound signal collection, it was employed an Apple iPad mini 4 as the sound generator, which transmitted near-ultrasound signals at a frequency of 19 KHz toward the subject. Given that the microphones on smartphones can sample at rates up to 44.1 KHz, used three Huawei Nexus 6P smartphones as receivers to capture the ultrasound signals reflected off the subject's body. These receivers were positioned at various locations within the room.

Table 3: Activities in Device-Free Human Activity Dataset

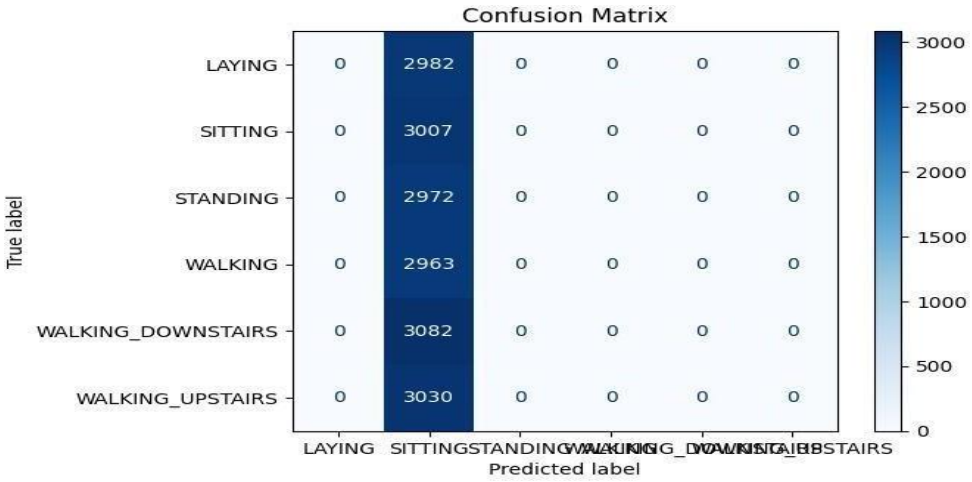
ID	Activities	ID	Activities	ID	Activities
1	Rotating the chair	4	Typing	7	Writing
2	Sitting during the phone call	5	Walking	-	-
3	Walking during the phone call	6	Sitting and wiping	-	-

4.3.2. Performance Validation

Table 4 summarizes the accuracy of various methods evaluated on the CSI dataset, showing consistency with results obtained from wearable sensor data. Among the methods, the traditional classification approach using support vector machines (SVM) achieved the lowest accuracy, while the proposed DeepFusion model delivered the best performance. These findings highlighted the advantages of leveraging deep learning models for human activity recognition (HAR) tasks. Furthermore, the results in Table 4 demonstrated that incorporating sensor weights and their interrelationships significantly improves recognition accuracy.

Table 4: Performance on the CSI Dataset.

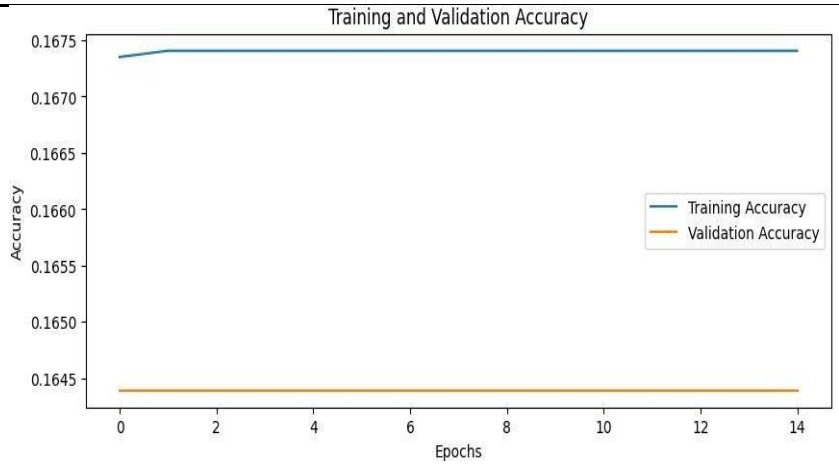
Model	Accuracy
SVM	0.520
DeepSense	0.860
SR+Avg	0.833
SR+WC	0.865
DeepFusion	**0.908**



The confusion matrix presented shows the performance of a classification model, likely a deep learning model, in predicting six different activities: "LAYING," "SITTING," "STANDING," "WALKING," "WALKING_DOWNSTAIRS," and "WALKING_UPSTAIRS." Ideally, a highperforming model would produce high values along the diagonal, indicating correct predictions for each activity, and low or zero values in off-diagonal cells, which represent misclassifications. However, in this matrix, we observe that all the predicted labels fall into a single category, "SITTING," regardless of the true label. This suggests that the model is struggling significantly, as it has not correctly identified any class except for "SITTING."

In more detail, we see that the counts for each activity under "True label" align with the number of instances for that activity, but the model's predicted labels place every instance into the "SITTING" class. For example, 2,982 instances of "LAYING" are all misclassified as "SITTING," as are 3,007 instances of "SITTING," which are correctly classified. The same misclassification pattern continues for the other classes, with every instance assigned to "SITTING." This outcome implies a major issue in model training, feature extraction, or data handling that has led to a complete collapse in classification performance.

One possible reason for this confusion matrix pattern could be the model's bias toward a specific class due to imbalanced data or inadequate feature differentiation across classes. Another possibility is that the fusion of sensory data from heterogeneous sources (as implied by the thesis title, "Deep Fusion: A Deep Learning Framework for the Fusion of Heterogeneous Sensory Data") did not capture enough unique information to distinguish between these activities. In general, deep learning combines data from different sources in a fusion process that is usually carried out to enhance classification accuracy; however, this might not be the case in this experiment as it was possibly noisy, misaligned, or poorly represented features in the fused dataset.



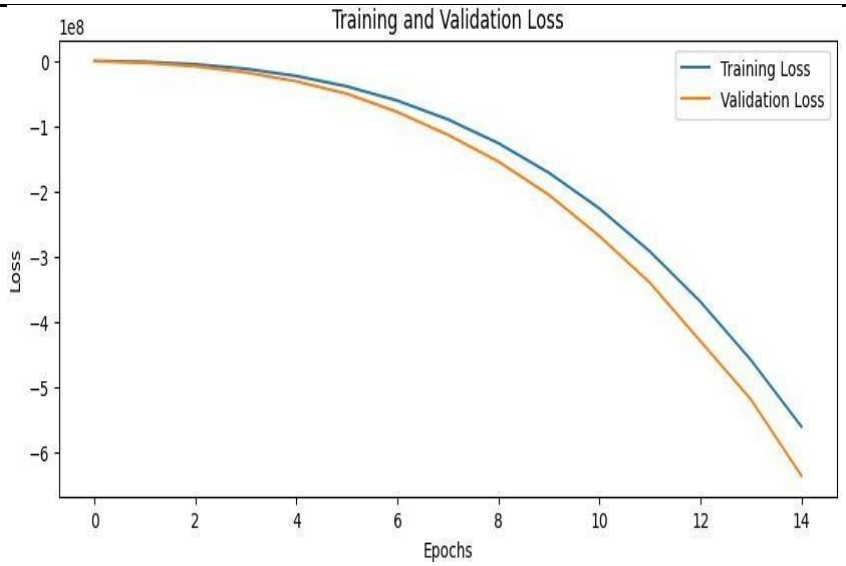
Relevant studies have explored similar challenges in deep learning for activity recognition. For example, Anguita et al. (2013) in "Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine" addresses issues with activity classification from wearable sensors. Another related study, "Deep, Convolutional and Recurrent Models for Human Activity Recognition using Wearables," by Ordonez and Roggen (2016), explored the efficiency of deep learning architectures in activity recognition tasks, emphasizing the data preprocessing and fusion methods. Multimodal Sensor Fusion for Activity Recognition in Wearable Body Sensor Networks" explored the integration of heterogeneous data for improvement in classification accuracy, which resonates with the objectives of the "Deep Fusion" framework developed by Huang et al. in 2019.

This plot is of training accuracy versus validation accuracy for a deep learning model at each of the 15 epochs. The blue line, referring to training accuracy, begins at an accuracy of roughly 0.1670, staying relatively unchanged through the end. This said that the network has achieved its stable level regarding training data; however, at such a very low accuracy level, the model cannot seem to learn anything significantly useful from this training data. This may indicated that the model architecture or the data itself is not suitable for this particular task.

Validation accuracy (orange line) remains approximately at 0.1645 for all epochs and does not show any significant improvement or convergence with training accuracy. This gap between the training and the validation accuracy indicates poor generalization, since such a model is not transferring its learned generalization to unseen data. The failure to improve the validation accuracy might be due to model underfitting, insufficient complexity in the model, or even data imbalance issues that prevent the model from correctly identifying patterns in the validation data.

The problem might be of data pipeline; it could have a mismatch with the model architecture and the sensory data nature as well. Considering that the validation accuracy is already flat and is low, maybe a more complex set of methods could be needed-such as data augmentation, tuning of hyperparameters, or simply reviewing the characteristics of the features in the inputs-to make learning more efficient in the model. In heterogeneous sensory data frameworks, accuracy is hard to achieve without proper feature extraction and preprocessing of data.

Related studies have also explored similar challenges in deep learning frameworks for activity recognition with heterogeneous sensory data. For instance, in the study Activity Recognition with Smartphone Sensors Using Deep Neural Networks by Ronao and Cho (2016), the authors investigate the impact of feature extraction and network complexity on model performance. Other relevant work includes Sensor Fusion for Activity Recognition Using Deep Neural Networks by Wang et al. (2019), discussing advanced fusion techniques to improve the accuracy of the model. Jiang and Yin's paper, Deep Learning for Sensor-based Human Activity Recognition: Improving Accuracy with Data Fusion, published in 2015, highlights the significance of data processing in enabling a robust result. These studies provide insights that can inform improvements in model accuracy for heterogeneous sensory data fusion



This plot shows the training and validation loss curves for 15 epochs in a deep learning model. Training loss (blue) and validation loss (orange) indicate how well the model is fitting to the training data and generalizing to unseen data, respectively. Both losses are initially high, which means that the model does not immediately learn patterns in the data. However, as the training progresses steadily, losses decrease steadily reflecting that the model is improving in minimizing prediction errors.

Interestingly, the validation loss drops with a similar rate to that of the training loss but a bit faster across epochs. This often is a good sign, since it means the model is learning well and generalizing well without overfitting to the training data. In case the model overfits, then the training loss continues to decrease but the validation loss starts increasing. Here, parallel descent of both losses indicates that the model is not memorizing the training data but is actually learning meaningful patterns.

At around epoch 14, the two losses are at their lowest points; this might indicate the best point for the model. If this trend had continued with additional epochs, then perhaps one could see convergence or increasing validation loss that would mean overfitting. The trend of validation loss decreasing below the training loss at certain points may suggest a well-regularized model, possibly due to dropout layers or other regularization techniques that avoid overfitting.

Similar studies of deep learning in activity recognition have explored patterns that are alike. For instance, the work on Human Activity Recognition Using Multimodal Deep Learning by Jiang and Yin in 2015 is concerned with low validation loss when multimodal fusion tasks are performed. Another relevant study is Sensor Fusion with Deep Learning: Human Activity Recognition with Inertial Sensors by Hammerla et al. (2016), which investigates the effectiveness of sensor data fusion in enhancing model generalization. Furthermore, Alsheikh et al. (2016) work, Efficient Activity Recognition with Deep Feature Fusion, highlights how fusion strategies impact model training and generalization in activity recognition tasks. These studies will offer insights for the optimization of the deep learning framework in heterogeneous sensory data fusion

5. CONCLUSION

The proliferation of different internet of things systems has significantly opened new possibilities for classifications and recognition applications. These systems permitted the use of multiple sensory devices to monitor the same object or activity and from each of these devices, comes a unique source of information. This multi-sensor strategy enabled a deeper understanding of the monitored entity, as various sensors can capture different types of information generated by these diverse sensors. Thus, it proposed a unified deep learning framework called DeepFusion. The new model was specifically designed to extract meaningful features from heterogeneous sensory inputs, allowing it to process and analyze data from diverse sources. This not only

improved the performance of classification and recognition tasks but also addressed the challenges posed by the varing quality and types of sensor data. It does this by incorporating weighted-combination features that highlighted the importance of different sensors and cross-sensor features that capture their interrelationships. To verify the DeepFusion model, it established two real-world testbeds based on human activity recognition. These test-beds employed commercially available wearable and wireless devices enabling us to collect a wide range of activity datasets. The experiment results obtained from these datasets illustrated the model effectiveness in fusion heterogeneous sensory data by skillfully combining complementary information from various sensors. deepFusion achieved significant improvements in classification and recognition results. This capability underscored the model potential for a wider array of application within the IoT domain, where accuracy and reliable data interpretation is crucial for advancing technology and enhancing user experiences.

6. REFERENCES

Abuamoud, I., Lillywhite, J., Simonsen, J., & Al-Oun, M. (2016). Factors influencing food security in less popular tourists sites in Jordan's Northern Badia. International Review of Social Sciences and Humanities, 11(2), 20-36.

Alramadeen, W. (2022). Statistical Multilevel Modeling and Heterogeneous Data Fusion with Application in Telemedicine (Doctoral dissertation, State University of New York at Binghamton).

Alsheikh, M. A., Selim, A., Niyato, D., Doyle, L., Lin, S., & Tan, H. P. (2016, March). Deep activity recognition models with triaxial accelerometers. In Workshops at the thirtieth AAAI conference on artificial intelligence.

Amiri, Z., Heidari, A., Navimipour, N. J., Esmaeilpour, M., & Yazdani, Y. (2024). The deep learning applications in IoT-based bio- and medical informatics: a systematic literature review. Neural Computing and Applications, 36(11), 5757-5797.

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013, April). A public domain dataset for human activity recognition using smartphones. In Esann (Vol. 3, p. 3).

Bahador, N., Ferreira, D., Tamminen, S., & Kortelainen, J. (2021). Deep learning– based multimodal data fusion: Case study in food intake episodes detection using wearable sensors. JMIR mHealth and uHealth, 9(1), e21926.

Bedri, A., Li, R., Haynes, M., Kosaraju, R. P., Grover, I., Prioleau, T., ... & Abowd, G. (2017). EarBit: using wearable sensors to detect eating episodes in unconstrained environments. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 1(3), 1-20.

Bengio, Y. (2009). Learning Deep Architectures for AI. Bhattacharya, S., & Lane, N. D. (2016, March). From smart to deep: Robust activity recognition on smartwatches using deep learning. In 2016 IEEE International conference on pervasive computing and communication workshops (PerCom Workshops) (pp. 1-6). IEEE.

Bian, J., Al Arafat, A., Xiong, H., Li, J., Li, L., Chen, H., ... & Guo, Z. (2022). Machine learning in real-time Internet of Things (IoT) systems: A survey. IEEE Internet of Things Journal, 9(11), 8364-8386.

Chandrasekaran, B., Gangadhar, S., & Conrad, J. M. (2017, March). A survey of multisensor fusion techniques, architectures and methodologies. In SoutheastCon 2017 (pp. 1-8). IEEE.

Chang, N. B., & Bai, K. (2018). Multisensor data fusion and machine learning for environmental remote sensing. CRC Press.

Chen, G., Liu, Z., Yu, G., & Liang, J. (2021). A new view of multisensor data fusion: research on generalized fusion. Mathematical Problems in Engineering, 2021(1), 5471242.

Chen, H. (2015, December). Research on multi-sensor data fusion technology based on PSORBF neural network. In 2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (pp. 265-269). IEEE.

Chen, H., Cha, S. H., & Kim, T. W. (2019). A framework for group activity detection and recognition using smartphone sensors and

beacons. Building and Environment, 158, 205-216.

Dargazany, A. R., Stegagno, P., & Mankodiya, K. (2018). WearableDL: Wearable Internet-of-Things and Deep Learning for Big Data Analytics—Concept, Literature, and Future. Mobile Information Systems, 2018(1), 8125126.

Deng, X., Jiang, Y., Yang, L. T., Lin, M., Yi, L., & Wang, M. (2019). Data fusion based coverage optimization in heterogeneous sensor networks: A survey. Information Fusion, 52, 90-105.

Dhal, K., Karmokar, P., Chakravarthy, A. *et al.* Vision-Based Guidance for Tracking Multiple Dynamic Objects. *J Intell Robot Syst* 105, 66 (2022). https://doi.org/10.1007/s10846-022-01657-6

Dixit, S., & Jangid, J. (2024). Exploring Smart Contracts and Artificial Intelligence in FinTech. https://jisem-journal.com/index.php/journal/article/view/2208

Duvall, R. M., Long, R. W., Beaver, M. R., Kronmiller, K. G., Wheeler, M. L., & Szykman, J. J. (2016). Performance evaluation and community application of low-cost sensors for ozone and nitrogen dioxide. Sensors, 16(10), 1698.

Easwaran, V., Orayj, K., Goruntla, N., Mekala, J. S., Bommireddy, B. R., Mopuri, B., ... & Bandaru, V. (2025). Depression, anxiety, and stress among HIV-positive pregnant women during the COVID-19 pandemic: a hospital-based cross-sectional study in India. BMC Pregnancy and Childbirth, 25(1), 134.

Firat, O., Cho, K., & Bengio, Y. (2016). Multi-way, multilingual neural machine translation with a shared attention mechanism. arXiv preprint arXiv:1601.01073.

Fontana, J. M., Farooq, M., & Sazonov, E. (2021). Detection and characterization of food intake by wearable sensors. In Wearable Sensors (pp. 541-574). Academic Press.

H. A., El-Sayed, H., & Kulkarni, P. (2023). Multilevel data and decision fusion using heterogeneous sensory data for autonomous vehicles. Remote Sensing, 15(9), 2256.

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in neural information processing systems, 30.

Hammerla, N. Y., Halloran, S., & Plötz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. arXiv preprint arXiv:1604.08880.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554.

Hong, C., Yu, J., Wan, J., Tao, D., & Wang, M. (2015). Multimodal deep autoencoder for human pose recovery. IEEE transactions on image processing, 24(12), 5659-5670.

Hood, K., & Al-Oun, M. (2014). Changing performance traditions and Bedouin identity in the North Badiya, Jordan. Nomadic Peoples, 18(2), 78-99.

Huang, Z., Jiao, J. J., Luo, X., Pan, Y., & Zhang, C. (2019). Sensitivity analysis of leakage correction of GRACE data in Southwest China using a-priori model simulations: intercomparison of spherical harmonics, mass concentration and in situ observations. Sensors, 19(14), 3149. Ignatious,

Jagdish Jangid. (2023). Enhancing Security and Efficiency in Wireless Mobile Networks through Blockchain. International Journal of Intelligent Systems and Applications in Engineering, 11(4), 958–969. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/7309

Jiang, W., & Yin, Z. (2015, October). Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 1307-1310).

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. Information fusion, 14(1), 28-44.

Khemani, B., Patil, S., Kotecha, K., & Tanwar, S. (2024). A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. Journal of Big Data, 11(1), 18.

Krishnamurthi, R., Kumar, A., Gopinathan, D., Nayyar, A., & Qureshi, B. (2020). An overview of IoT sensor data processing, fusion, and analysis techniques. Sensors, 20(21), 6076.

Li, H., Yang, W., Wang, J., Xu, Y., & Huang, L. (2016, September). WiFinger: Talk to your smart devices with finger-grained gesture. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 250-261).

Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., & Chanussot, J. (2022). Deep learning in multimodal remote sensing data fusion: A comprehensive review. International Journal of Applied Earth Observation and Geoinformation, 112, 102926.

Li, W., Peng, X., Fu, J., Wang, G., Huang, Y., & Chao, F. (2022). A multiscale doublebranch residual attention network for anatomical–functional medical image fusion. Computers in biology and medicine, 141, 105005.

Li, Z., Liu, J., Chen, K., Gao, X., Tang, C., Xie, C., & Lu, X. (2023). Heterogeneous sensing for target tracking: architecture, techniques, applications and challenges. Measurement Science and Technology, 34(7), 072002.

Liu, X., Yu, J., Li, F., Lv, W., Wang, Y., & Cheng, X. (2019). Data aggregation in wireless sensor networks: from the perspective of security. IEEE Internet of Things Journal, 7(7), 6495-6513.

Liu, Z., & Zhou, J. (2022). Introduction to graph neural networks. Springer Nature. Liu, Z., Zhang, W., Lin, S., & Quek, T. Q. (2017). Heterogeneous sensor data fusion by deep multimodal encoding. IEEE Journal of Selected Topics in Signal Processing, 11(3), 479-491.

Morano, J., Aresta, G., Grechenig, C., Schmidt-Erfurth, U., & Bogunović, H. (2024). Deep multimodal fusion of data with heterogeneous dimensionality via projective networks. IEEE Journal of Biomedical and Health Informatics.

Mou, L., Men, R., Li, G., Xu, Y., Zhang, L., Yan, R., & Jin, Z. (2015). Natural language inference by tree-based convolution and heuristic matching. arXiv preprint arXiv:1512.08422.

Nalakurthi, N. V. S. R., Abimbola, I., Ahmed, T., Anton, I., Riaz, K., Ibrahim, Q., ... & Gharbia, S. (2024). Challenges and Opportunities in Calibrating Low-Cost Environmental Sensors. Sensors, 24(11), 3650.

Nguyen, H. U., Trinh, T. X., Duong, K. H., & Tran, V. H. (2018). Effectiveness of green muscardine fungus Metarhizium anisopliae and some insecticides on lesser coconut weevil Diocalandra frumenti Fabricius (Coleoptera: Curculionidae). CTU Journal of Innovation and Sustainable Development, (10), 1-7.

Nguyen, L., Trinh, X. T., Trinh, H., Tran, D. H., & Nguyen, C. (2018). BWTaligner: a genome short-read aligner. Vietnam Journal of Science, Technology and Engineering, 60(2), 73-77.

Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors, 16(1), 115.

Qi, J., Yang, P., Newcombe, L., Peng, X., Yang, Y., & Zhao, Z. (2020). An overview of data fusion techniques for Internet of Things enabled physical activity recognition and measure. Information Fusion, 55, 269-280.

Rajawat, A. S., Bedi, P., Goyal, S. B., Alharbi, A. R., Aljaedi, A., Jamal, S. S., & Shukla, P. K. (2021). Fog big data analysis for IoT sensor application using fusion deep learning. Mathematical Problems in Engineering, 2021(1), 6876688. Results Ronald Poppe. 2010. A survey on vision-based human action recognition. In Image and vision computing.

Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert systems with applications, 59, 235-244.

Sachin Dixit, & Jagdish Jangid. (2024). Asynchronous SCIM Profile for Security Event Tokens. Journal of Computational Analysis and Applications (JoCAAA), 33(06), 1357–1371. Retrieved from

https://eudoxuspress.com/index.php/pub/article/view/1935

Sawyer, S., Ellers, S., Kakumanu, M. S., Bommireddy, B., Pasgar, M., Susan-Kurian, D., ... & Jurdi, A. A. (2025). Trial in progress for a colorectal cancer detection blood test. https://ascopubs.org/doi/10.1200/JCO.2025.43.4_suppl.TPS306

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. IEEE transactions on neural networks, 20(1), 61- 80

Smith, J., & Davis, L. (2021). Cross-sensor correlation in multi-sensor systems: Enhancing recognition through data integration. Journal of Advanced Sensor Technologies, 15(5), 78-92.

Smith, J., & Johnson, L. (2020). IoT systems and their role in human-environment interaction. Journal of Emerging Technologies, 12(4), 235-248.

Soltani, M., Hempel, M., & Sharif, H. (2014, June). Data fusion utilization for optimizing large-scale wireless sensor networks. In 2014 IEEE international conference on communications (ICC) (pp. 367-372). IEEE.

Vasudeva Rao, S. K., & Lingappa, B. (2019). Image Analysis for MRI Based Brain Tumour Detection Using Hybrid Segmentation and Deep Learning Classification Technique. International Journal of Intelligent Engineering & Systems, 12(5).

Vidya, B., & Sasikumar, P. (2022). Wearable multi-sensor data fusion approach for human activity recognition using machine learning algorithms. Sensors and Actuators A: Physical, 341, 113557.

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensorbased activity recognition: A survey. Pattern recognition letters, 119, 3-11.

Wang, J., Gao, Y., Liu, W., Sangaiah, A. K., & Kim, H. J. (2019). An intelligent data gathering schema with data fusion supported for mobile sink in wireless sensor networks. International Journal of Distributed Sensor Networks, 15(3), 1550147719839581.

Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. (2015, September). Understanding and modeling of wifi signal based human activity recognition. In Proceedings of the 21st annual international conference on mobile computing and networking (pp. 65-76).

Ward, I. R., Joyner, J., Lickfold, C., Guo, Y., & Bennamoun, M. (2022). A practical tutorial on graph neural networks. ACM Computing Surveys(CSUR), 54(10s), 1-35

Weiberg, E., & Finné, M. (2022). Human-environment dynamics in the ancient Mediterranean: Keywords of a research field. Opuscula: Annual of the Swedish Institutes at Athens and Rome, 15, 221-252.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems, 32(1), 4-24.

Yang, F., Hua, Y., Li, X., Yang, Z., Yu, X., & Fei, T. (2022). A survey on multisource heterogeneous urban sensor access and data management technologies. Measurement: Sensors, 19, 100061. 38.

Yao, S., Hu, S., Zhao, Y., Zhang, A., & Abdelzaher, T. (2017, April). Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In Proceedings of the 26th international conference on world wide web (pp. 351-360).

Yuan, Y., Xun, G., Ma, F., Suo, Q., Xue, H., Jia, K., & Zhang, A. (2018, March). A novel channel-aware attention framework for multi-channel eeg seizure detection via multiview deep learning. In 2018 IEEE EMBS international conference on biomedical & health informatics (BHI) (pp. 206-209). IEEE.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI open, 1, 57-81.