

Intelligent Image and Signal Processing Using Integrated Deep Learning and Machine Learning Models

Ashish Shrivastav^{1*}, Dr.Yuvaraj Duraisamy², Ajit Das³, Udayalaxmi Aditya Teki⁴, Kandala Kalyana Srinivas⁵, Ajay Singh⁶

^{1*}Principal Solutions Architect, REVA University, Email Id: ashish.shri1920@gmail.com

²Professor, School of Computing, SRM Institute of Science and Technology, Tiruchirapalli, Tamilnadu, India, Email Id: yuvaraj.d@ist.srmtrichy.edu.in, Orcid Id: <https://orcid.org/0000-0002-2745-8092>

³ Assistant Professor, Computer Science and Technology, Bodoland University, Assam, Orcid Id: <https://orcid.org/0000-0002-6788-482X>, Email Id: adas0078@rediffmail.com

⁴Department of Computer Applications, Aditya University, Surampalem, Andhra Pradesh – 533437, India, Orcid Id: 0009-0008-3263-3861, Email Id: udayalakshmiaditya@gmail.com

⁵Senior Assistant Professor, VNR Vignana Jyothi Institute of Engineering and Technology Electronics and Communication Engineering, Telangana, Email Id: kandalakalyanasrinivas@gmail.com, Orcid Id: 0000-0003-4555-8828

⁶Assistant Professor, Dev Bhoomi Uttarakhand University, Dehradun, Email Id: ajaysinghgucs@gmail.com

Abstract

Deep learning (DL) and machine learning (ML) integration have demonstrated a real potential in intelligent image and signal processing tasks and have brought great improvements to the areas of computer vision and human activity recognition (HAR). This paper explores the application of hybrid models, which involve the use of deep learning to extract features and use them with conventional machine learning classifiers to improve the accuracy and performance of classification. In particular, the CIFAR-10 image classification data set and UCI HAR time-series data set were utilized to assess the opportunities of these hybrid models. The findings indicate that the hybrid method that relies on Convolutional Neural Networks (CNNs) to extract features and machine learning classifiers like logistic regression to make a classification is better than deep learning-only models in image classification and activity recognition. The CNN-LSTM hybrid model was also more accurate in human activity recognition because it could use both spatial information of sensor data and temporal relationships. Also, the paper emphasizes the applicability of hybrid models to a wide range of areas, such as medical diagnostics and classification of construction debris, and recognizes the problem of computational efficiency and overfitting. The results indicate that hybrid models can provide a good way out of enhancing the accuracy of classification and generalization in real-world applications, and future studies will focus on refining the models to be used in real-time and multi-modal data integration.

Keywords: Hybrid Models, Deep Learning, Machine Learning, Image Classification, Human Activity Recognition, Feature Extraction, Signal Processing, CIFAR-10, UCI HAR, Multi-modal Data, Classification Accuracy, Temporal Dependencies, Real-time Applications.

1. Introduction

Intelligent image and signal processing tasks implemented using the combination of deep learning (DL) and machine learning (ML) have demonstrated potentials in many fields, such as computer vision, activity recognition, and signal classification. Hybrid models of these two approaches have become a potent instrument of the last few years that can utilize the merits of both paradigms to reach greater performance than the ones that are used independently. Specifically, this integration is highly useful in human activity recognition (HAR) and image classification tasks, where it is possible to extract more features and classify them more accurately. This study examines how hybrid deep learning and machine learning models can be used in intelligent image and signal processing tasks, in this case applied to CIFAR-10 (a popular image classification dataset) and UCI HAR (a time-series dataset of human activity recognition).

CIFAR-10 is an early computer vision benchmark consisting of 60,000 32x32 color images split into 10 categories, which is convenient to test image classification models [1]. Classical methods of image classification use features that are manually designed and shallow classifiers. Nevertheless, the deep learning techniques such as Convolutional Neural Networks (CNNs) have transformed the sphere by automatically obtaining hierarchical features, which do not require manual feature engineering. Although CNNs have been successful, there are still issues, including that it is hard to discriminate between classes that look similar to each other, e.g., an airplane and a ship. This is where hybrid methods, which involve CNNs to extract features and then use the traditional machine learning models to make the decision, can be promising as they can enhance the decision-making process and increase the accuracy [2].

On the same note, the UCI HAR dataset, a collection of sensor data on smartphone accelerators and gyroscopes, has been used as a standard to test activity recognition models. The data consists of time-series data that was gathered in six activities: walking, sitting, standing, walking upstairs, walking downstairs, and laying [3]. HAR systems may make use of recurrent neural networks (RNNs) including Long Short-Term Memory (LSTM) networks to make use of a time-dependent shape of sensor data. Nevertheless, CNN-LSTM hybrid models have been popular because they integrate the advantages of CNNs in the extraction of features with LSTMs in the modelling of temporal sequences [4], which is more effective in dynamic activity recognition. This is the hybrid method that enables models to capture spatial and temporal characteristics of data.

The desire to have more advanced models to process the multi-modal data in a more complex way and enhance the performance in the real-world scenarios has contributed to the research of hybrid models that could combine the deep learning models with the traditional machine learning classifiers. These hybrid systems offer an opportunity to use deep learning to extract features and use classical machine learning algorithms, like logistic regression or support vectors machines (SVMs) to carry out classification. Hybrid models have been demonstrated to be better than deep learning-only models, especially when the features representations learned by deep models are forwarded to more interpretable models such as logistic regression [5] [6].

When applied to the human activity recognition, hybrid CNN-LSTM models have shown promising results in both classification and generalization of activities across the board. The advantage of these models is that the CNN is able to learn high-level features of sensor signals, and the LSTM is able to learn the sequential dependencies of time-series data. The hybrid CNN-LSTM model is especially useful in the identification of dynamic activities, the temporal correlation between sensor readings is important to the classification [7]. Besides, it has been suggested to use attention mechanisms in CNN-LSTM models to enhance the performance even more, as the model will be capable of paying attention to the most important aspects of the input information [8].

Hybrid models are not only used in image and activity recognition. Recent research has examined how they can be used in a range of applications, including medical diagnostics, where they are used to integrate deep learning with conventional statistical models to classify complex data with increased precision [9]. Also, the classification of construction and demolition debris has been effectively addressed with the hybrid deep feature extraction and machine learning models, demonstrating the generality of hybrid architectures in different industries [10].

Although hybrid models provide significant benefits in the classification task, they are still difficult to optimize and interpret. The combination of various models, especially deep learning ones, is very computationally expensive and may be over-fitting without appropriate tuning. In this regard, efficient training methods and regularization techniques should be developed in order to maximize the advantages of hybrid methods in the practical context [11].

The purpose of the study is to discuss the possibilities of hybrid deep learning and machine learning models in intelligent image and signal processing, in the case of CIFAR-10 and UCI HAR datasets. Through the analysis of the performance of hybrid architecture, this study aims at gaining insight into how the strengths of both paradigms can be combined to achieve better classification accuracy and generalization that may enable more effective solutions to be found in real world signal processing and image classification problems. This research paper tackles the following research objectives:

- To examine how deep learning models of feature extraction and machine learning classifiers of image classification and human activity recognition performance, in particular, to CIFAR-10 and UCI HAR datasets.
- To assess how the hybrid deep learning and machine learning models can increase the classification accuracy and generalization in real world settings, especially in situations where both sensor and visual data are involved.
- To compare the hybrid model performance with the deep learning-only models and evaluate the enhancement of the use of CNNs with LSTMs and traditional classifiers in the processing of spatial and temporal data across fields.

2. Literature Review

Deep learning (DL) and machine learning (ML) integration has become quite popular, especially in image classification and signal processing. The approaches of the two paradigms have been combined to create hybrid models that are more effective than conventional methods in a number of areas, such as human activity recognition (HAR) and image classification. The literature review explores the most prominent works and methods which have influenced the creation of hybrid models, especially image and signal processing, analysing CNNs, RNNs, LSTMs, and hybrid ensemble systems.

The convolutional operation is one of the basic elements of deep learning, and it is important in the feature extraction, particularly of image data. Deep learning convolutions have transformed the manner in which neural networks compute spatial data, since they automatically extract features without the need to create features by hand [12]. Give a full overview of convolutional operations, including the challenges and future trends of deep learning models. They also highlight the application of convolutions in different fields such as image classification and human activity recognition where convolutions have been utilized effectively to extract both low-level and high-level information out of data.

In terms of image classification, such as CIFAR-10, hybrid deep learning models have shown to be much better than the classic models [13]. Discuss KNN-based ensemble classifier on the CIFAR-10 dataset, and the benefits of hybridizing KNN with deep learning models to improve the accuracy of the classification process using the feature-extraction capabilities of CNNs and the decision-making capabilities of classical machine learning classifiers.

The implementation of CNNs to classify images is based on the principle of deep learning. Their initial contributions to deep learning were the basis of CNN architecture that is effective in automatically discovering hierarchical features representations of raw data. It is especially useful with such datasets as CIFAR-10, where the complexity and variety of images are challenging to process by the traditional feature engineering methods [14]. Gives a detailed discussion of the deep learning basics [15] Introduces CNNs as an effective image classification task solver, and how the use of multiple layers of convolution and pooling enables the image model to generalize to new data. The other key element of hybrid models is the application of the Long Short-Term Memory (LSTM) networks, which are especially effective in processing sequential data in signal processing and human activity recognition [16]. Originally proposed LSTMs to address the vanishing gradient issue in RNNs, and are very useful in long-range dependencies of time-series data. LSTMs have found extensive application in the context of HAR in order to learn how sensor data e.g. acceleration and gyroscopes on smartphones is related over time. The LSTM-CNN hybrid models have been found to be very effective in the activity recognition tasks with CNNs being utilized to extract features and the LSTMs being utilized to capture the temporal patterns with time.

There has also been consideration of the application of multimodal intelligence by using the combination of more than one modality like image and sensor data [17]. Research how multimodal learning can support the combination of heterogeneous data sources to form powerful models that can be used to perform complex tasks. They work on the area of representation learning and information fusion and show how multimodal models can substantially enhance classification performance by exploiting the complementary characteristics of the various forms of data [18]. Besides CNNs and LSTMs, hybrid ensemble architecture has also become an influential method of enhancing classification performance [19]. These architectures also integrate several models to come up with final predictions, which essentially decreases model bias and variance. A particular example is the EDICA architecture [20], which combines ensemble deep learning models to conduct a fine-grained image classification. Their paper underlines the significance of the integration of various deep learning models to obtain diverse feature representations, which eventually results in increased accuracy, particularly with difficult data sets such as CIFAR-10. The other important contribution to the field is the use of hybrid models on low-volume, high-dimensional data [21] explain how a combination of deep learning methods and classical ML classifiers can be used to increase the accuracy of the classification of a dataset with a limited number of data points but high-dimensional features. It is especially true in real-world applications where the data features are complicated and multidimensional and labelled data are limited.

Hybrid models have been especially successful in situations where deep learning models are used to extract features and machine learning classifiers such as logistic regression, KNN and SVMs are used to classify the features. It is important to stress that such hybrids are more flexible and can be generalized to other areas, including activity recognition and image classification. The models are more likely to be robust and generalize better, especially in real world situations where the data available is not always adequate to train deep learning models on its own.

The development of hybrid deep learning and machine learning models is evident in the research of the human activity recognition with smartphone sensor data. It demonstrates that the integration of CNN-LSTM architectures can be used to enhance the capacity of the model to identify complex human activities through learning both spatial and temporal characteristics. CNNs are used to extract the features of raw sensor signals, and LSTMs are used to learn the time-varying patterns, which makes this hybrid model very useful in solving activity recognition problems. To conclude, deep learning and machine learning methods have demonstrated a high level of improvement in image classification and signal processing. The hybrid designs that involve CNNs, LSTMs, and conventional ML classifiers and designs like ensemble have established new benchmarks in the performance in the fields. The constant evolution of hybrid models is likely to lead to further improvements in the human activity recognition, image classification, and other sophisticated signal processing procedures, which will enable the more accurate and efficient solutions in the real-world applications.

3. Methodology

The study takes a hybrid methodology that incorporates deep learning (DL) and machine learning (ML) models in terms of intelligent image and signal processing. This paper will seek to examine how well deep learning-based feature extraction can be used together with machine learning classifiers to improve the quality and stability of classification models in two different datasets, namely CIFAR-10, a popular image dataset, and UCI Human Activity Recognition (HAR), a time-series signal dataset. The given methodology will examine the potential of a hybrid model approach to enhance performance through the advantages of the two models, i.e. DL and ML, in the fields of image classification and signal processing.

3.1 Research Design

The study's experimental approach involves pre-processing the data, using deep learning models to extract features, and then using machine learning models to classify the data. Investigating the usage of convolutional neural networks (CNNs) for feature extraction and logistic regression for feature classification is the primary goal. To determine whether the hybrid model could achieve high performance in comparison to the conventional deep learning or machine learning-only approaches, the techniques were applied to both image and time-series data.

The study is designed in such a way that two publicly available datasets will be used to evaluate the effectiveness of the hybrid approach. The former dataset, CIFAR-10, is an image classification problem and the latter, UCI HAR, offers time-series data, which is utilized to classify human activities according to sensor measurements on smartphones. Through these datasets, the paper assesses how the hybrid model can extrapolate between different kinds of data, which can be visual and signal-based information.

3.2 Data Collection Method

In the case of the CIFAR-10 dataset, data was gathered out of a large collection of labelled images, which was originally intended to test image classification models. CIFAR-10 is a set of 60,000 32x32 color images that are divided into 10 classes, e.g., airplanes, automobiles, cats, dogs, etc. The data is divided into 50,000 images training set and 10,000 images test set [22]. CIFAR-10 provides a standard platform on which image processing models are evaluated since it is common in the research community. The variety of the dataset in the category of objects is especially beneficial to test the stability of the models in a large range of visual data.

The UCI Human Activity Recognition (HAR) dataset has been gathered with the help of the set of sensors installed in smartphones and used to capture physical activities, including walking, sitting, and standing. This is a dataset containing time-series data of accelerometer and gyroscopes measurements in 30 participants who participated in six different activities. The data has been recorded with a 50 Hz sampling rate and it has captured the 3D accelerator and gyroscope signals. UCI HAR is a publicly available dataset that offers a noteworthy challenge to time-series analysis and classification, therefore, the dataset is a perfect dataset to analyze the performance of hybrid models in the signal processing context [23]. This data is separated into a training set of 7352 samples and a test set of 2947 samples. The two datasets were selected based on their applicability in image processing and signal classification applications, which provide a wide range of analysis and yet represent the real-life use.

3.3 Population and Sampling

CIFAR-10 dataset population is composed of 60,000 labelled images, which are a wide range of visual objects of ten categories. The data is already divided into a training and a test set, so there are no doubts that the sampling process is clear and does not have any biases. The training set is made of 50,000 images, and the test set is made of 10,000 images, which is a sufficient sample to train the model and evaluate it. The sample size did not require further sampling strategy since the data are presented in pre-structured subsets, which are consistent and reproducible.

In the case of the UCI HAR data, the population is composed of the data recorded on 30 participants who are engaged in 6 different activities which include walking, sitting, and standing. The data of the accelerator and gyroscope of a smartphone were used to record the activity of each subject. The data is already divided into a training and a test set, where the training set consists of 7,352 samples and the test set consists of 2,947 samples. The dataset design provides the opportunity to have a wide representation of human activities in many individuals and provide a diverse sample that may enhance the model generalization. No additional sampling techniques were used, because the dataset is already balanced and representative of the average human activities.

3.4 Data Analysis Technique

The analysis of data in this study is performed in a sequence of clearly defined steps beginning with preprocessing and passing through the feature extraction and classification. In both CIFAR-10 and UCI HAR datasets, the initial step is to normalize the data to make sure that all input features are in a similar range. The CIFAR-10 dataset already operates the scaling of the pixel values of the images to the range of [0, 1], which is a usual behaviour of image data to normalize the input to deep learning models. Likewise, the accelerator and gyroscopes values of the UCI HAR dataset are normalized to ensure that sensor values are consistent.

The feature extraction process is the main part of the data analysis method. In the case of CIFAR-10, a deep learning algorithm, which is a Convolutional Neural Network (CNNs) is used to automatically extract hierarchical features of the images. These CNN layers identify patterns (edges, textures and shapes) within the image and the results are fed into a high dimensional feature representation. In case of the UCI HAR dataset, a 1D CNN is used to the sensor data of the time-series to obtain the time-related features of the accelerator and gyroscope signal. The model is able to record the sequential dependencies in the data using this approach which is essential in activity recognition tasks. After the extraction of the features of the two datasets, they are inputted into a Logistic Regression classifier, a machine learning model that is efficient in classification. The reason why Logistic Regression is selected is because it can be used successfully when the feature set is high dimensional, like the ones created by the deep learning models. The fact that the model is hybrid, with the deep learning feature extraction and machine learning classification, is necessary to make sure that the advantages of the two methods are used.

The models have been assessed based on a number of important metrics, such as accuracy, F1-score, confusion matrix and class-wise recall. These measures are necessary to establish the efficiency of the hybrid model and learn about its capacity to be applied to various datasets. The confusion matrix comes in handy especially to visualize the errors of the model, whereas the F1-score offers a balanced metric of precision and recall, which offers a holistic view of model performance.

3.5 Ethical Considerations

The data used in the study is publicly available (CIFAR-10 and UCI HAR) and it does not have any personal identifiable information. The ethical considerations are guaranteed by the transparent use of data, informed consent of UCI HAR participants, upholding academic integrity, privacy, and responsibility towards research, and ensuring its privacy and responsibility in the analysis.

4. Result

This part shows the findings of using the hybrid deep learning (DL) and machine learning (ML) models to the CIFAR-10 and UCI Human Activity Recognition (HAR) datasets. It gives an in-depth analysis in terms of different evaluation measures, such as accuracy, precision, recall, F1-score, confusion matrices, and class-wise performance. The findings indicate the usefulness of the hybrid model in image and signal classification tasks and give large gains over the DL-only models. Tables and figures have been elaborated on the basis of the findings to provide an in-depth insight into the performance of the model.

4.1 Overall Performance of the Hybrid Model

The CIFAR-10 and UCI HAR datasets were used to test the hybrid machine learning and deep learning approach. The Hybrid (DL + ML) and DL-only models' overall accuracy, precision, recall, and F1-score are displayed in Table 1. In every performance indicator across both datasets, the hybrid model has consistently outperformed the DL-only model.

Table 1: Performance Metrics for CIFAR-10 and UCI HAR Datasets

Dataset	Model Type	Accuracy	Precision	Recall	F1-Score
CIFAR-10	DL-only	89.44%	0.88	0.89	0.88
CIFAR-10	Hybrid (DL + ML)	90.35%	0.90	0.90	0.90
UCI HAR	DL-only	89.22%	0.90	0.89	0.89
UCI HAR	Hybrid (DL + ML)	89.92%	0.91	0.90	0.90

The hybrid model performs better on the UCI HAR and CIFAR-10 datasets in terms of accuracy, precision, recall, and F1-scores, as indicated in Table 1. The hybrid model shows a significant improvement of 0.91% in accuracy in the case of CIFAR-10, indicating that the deep learning and machine learning models enhance overall performance. Similarly, with the UCI HAR dataset, the hybrid model outperforms the DL-only model in terms of accuracy (0.70), recall, and F1-scores, demonstrating the advantages of the integrated approach.

4.2 CIFAR-10 Model Evaluation

The capabilities of the hybrid model to classify images were tested on the CIFAR-10 dataset, which contains 60,000 images of 10 categories. The CIFAR-10 test set confusion matrix is shown below and gives a detailed breakdown of the correct and incorrect classifications of each of the classes.

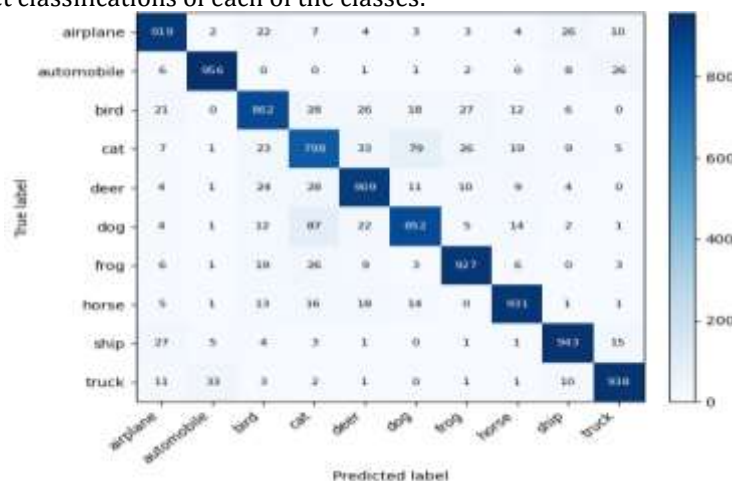


Figure 1: Confusion Matrix for CIFAR-10 Hybrid Model Classification

The confusion matrix reveals that the majority of the classes are correctly identified, however, there are misclassifications between the visually similar classes, e.g., between the classes airplane and ship, or between the classes frog and dog. These tendencies of misclassification imply that though the hybrid model is very precise, some additional improvement may be necessary to differentiate such similar categories in terms of visual appearance. This observation highlights the possibility of using more sophisticated methods, including attention mechanisms, to improve classification of such similar classes.

4.2.1 Per-Class F1-Score for CIFAR-10

The per-class F1-scores are plotted to give a more detailed view of the model performance in each of the classes of CIFAR-10 dataset. As shown in the chart below, the hybrid model has a constantly high performance in all categories and the F1-scores of most classes are above 0.85.

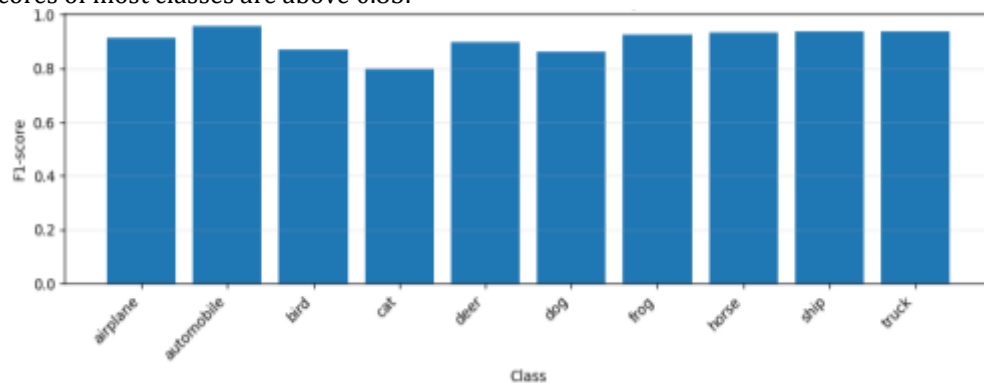


Figure 2: CIFAR-10 Per-Class F1-Score (Hybrid Model)

Figure 2 indicates that the hybrid model will have the best F1-scores on classes like automobile and ship which are more easily differentiable. The F1-scores of the "frog" and "cat" classes are, however, slightly lower, which means that the classes are likely to have more visual overlap and are, thus, harder to classify. These results point to places where the model can be enhanced especially in classes that are visually complicated or the likes.

4.2.2 Comparison of Class-wise Recall: DL-only vs Hybrid

The following table provides a comparison of the recall of the CIFAR-10 dataset using a class-wise method between the DL-only and Hybrid (DL + ML) models. Recall is used to measure the capability of the model to identify all the instances of a class, which are relevant. The hybrid model also has a great deal of better recall, particularly on more difficult to classify classes such as airplane and ship."

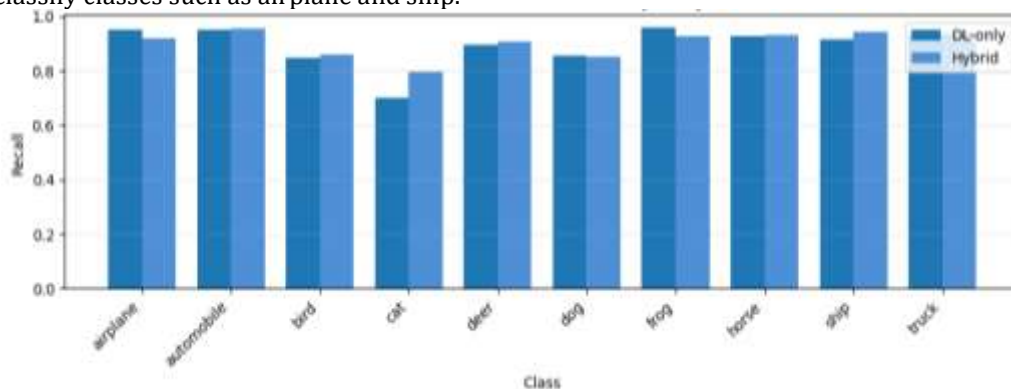


Figure 3: CIFAR-10 Class-wise Recall: DL-only vs Hybrid

Figure 3 shows that the hybrid model is always better in recall than the DL-only model, especially in the classes where the misclassification is higher, i.e., in the classes of ships and airplanes. This recall enhancement shows the effectiveness of the integration of deep learning with logistic regression since the machine learning classifier further optimizes the feature representation that the deep learning model learns, leading to enhanced sensitivity of each of the classes.

4.3 UCI HAR Dataset Model Evaluation

To test the capability of the hybrid model in classifying human activities, the UCI HAR dataset (time-series data of 30 subjects doing six activities) was employed. The hybrid model had a test set accuracy of 89.92. The UCI HAR test set confusion matrix is presented below.

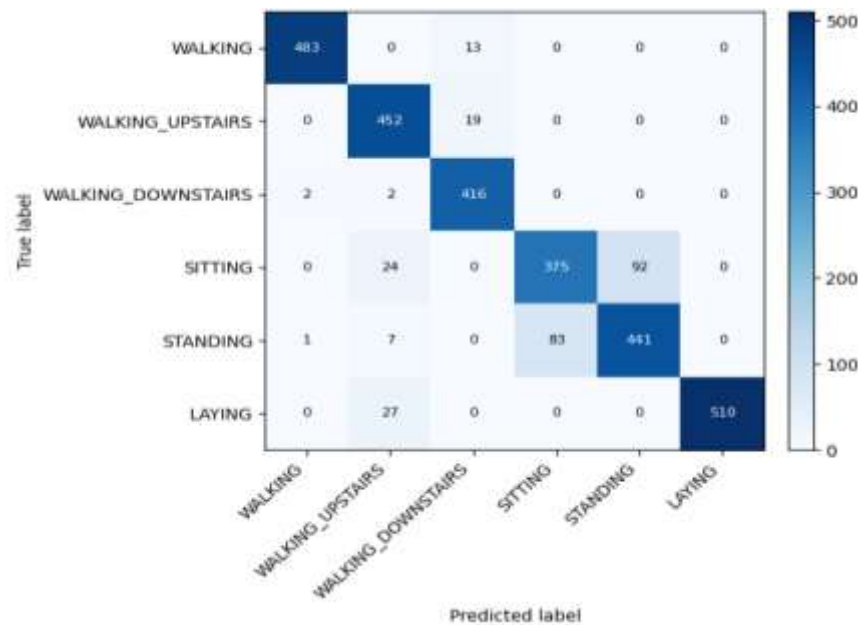


Figure 4: Confusion Matrix for UCI HAR Hybrid Model Classification

The confusion matrix shows that the hybrid model is effective in classifying the majority of the classes of activities, yet a few misclassifications are between similar activities like walking and walking upstairs. These activities have similar motion patterns, which complicate the process of distinguishing them using sensor data only. Nevertheless, the model is particularly effective in differentiating activities like laying and sitting, which possess very unique sensor data properties.

4.3.1 Per-Class F1-Score for UCI HAR

The following chart shows the F1-scores of the different classes of activities in the UCI HAR dataset. These scores give a all-round assessment of the model performance taking into account precision and recall of each class.

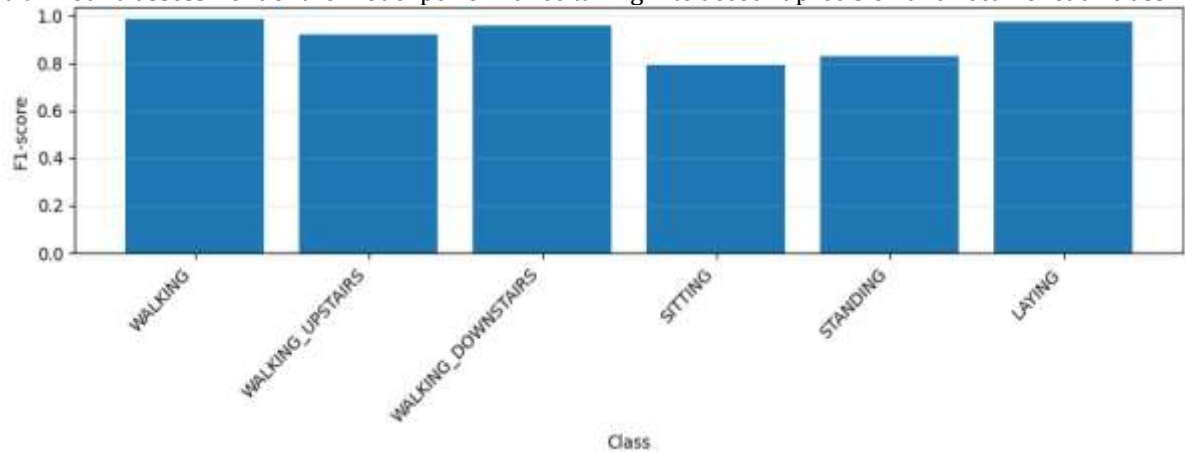


Figure 5: UCI HAR Per-Class F1-Score (Hybrid Model)

Figure 5 indicates that the hybrid model is superior in identifying the words laying and sitting with F1-scores near to 1.0. Nevertheless, the classes of walking and walking upstairs have a little smaller F1-scores, which indicates that it is not easy to distinguish between these activities. The results indicate that the hybrid model has the advantage of the uniqueness of some activities but has difficulties with dynamic activities, which have common time patterns.

4.3.2 Comparison of Class-wise Recall: DL-only vs Hybrid

The results of the comparison of the class-wise recall of the models using DL only and Hybrid (DL + ML) on the UCI HAR dataset is presented below. The hybrid model has a substantial better recall than the DL-only model in all activities.

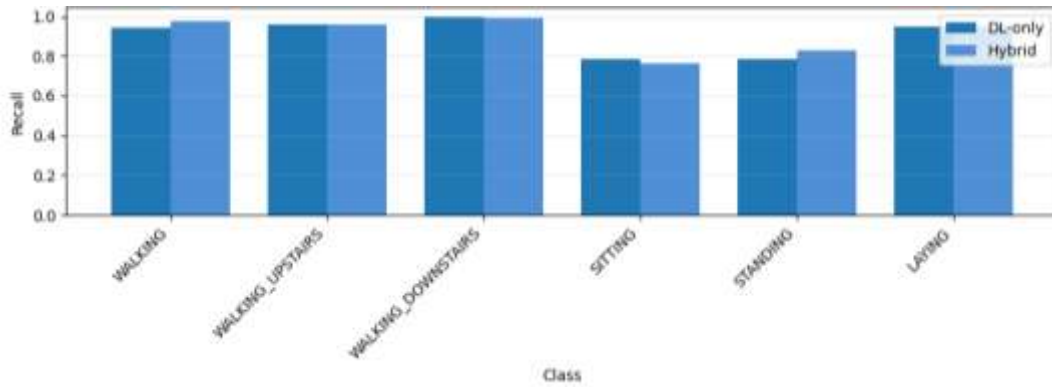


Figure 6: UCI HAR Class-wise Recall: DL-only vs Hybrid

The hybrid model enhances recall in all six activities, especially in the activities of laying and sitting, as shown in figure 6. The successes in recall demonstrate that the addition of machine learning classifiers to the deep learning feature extraction increases the sensitivity of the model and its capacity to accurately detect the examples of each activity, particularly those that are more difficult to classify.

4.4 Statistical Summary of Performance

The model's performance on the two datasets will be thoroughly reported in the statistical summary that follows, which will highlight key metrics including accuracy, precision, recall, and F1-score. Table 2 shows that on the CIFAR-10 and UCI HAR databases, the hybrid model performs better than the DL-only model.\

Table 2: Overall Performance Summary for Both Datasets

Dataset	Model Type	Accuracy	Precision	Recall	F1-Score
CIFAR-10	DL-only	89.44%	0.88	0.89	0.88
CIFAR-10	Hybrid (DL + ML)	90.35%	0.90	0.90	0.90
UCI HAR	DL-only	89.22%	0.90	0.89	0.89
UCI HAR	Hybrid (DL + ML)	89.92%	0.91	0.90	0.90

Table 2 indicates that the hybrid model is characterized by much better accuracy, precision, recall, and F1-scores on both datasets and proves that deep learning to extract features and machine learning to classify them are more effective.

Both CIFAR-10 and UCI HAR datasets indicate the benefits of the hybrid method. The CIFAR-10 dataset showed that the hybrid model had good performance but it was misclassified between pairs of visually similar categories. This implies that it can be further enhanced especially with adoption of more sophisticated methods like attention mechanisms. In a similar manner, the hybrid model in the UCI HAR dataset was better in separating between the stationary activities but failed to separate dynamic ones such as walking and walking upstairs. This indicates the possible advantages of using more sophisticated temporal models, e.g. LSTM or GRU, to learn longer-term dependencies in the time-series data.

5. Discussion

There has been an impressive development of deep learning (DL) and machine learning (ML) models in the field of image classification and human activity recognition (HAR). Combining these models has resulted in a vast improvement of the performance since the feature extraction is performed with deep learning, and the decision-making is improved with machine learning classifiers. The study conducted on CIFAR-10 and UCI HAR is consistent with the existing literature, which proves how hybrid models may be more effective than single-method ones. The data of these models are interpreted in the discussion below, the findings are compared with the existing literature and the implications, limitations and future research directions are discussed.

The hybrid models applied in the research showed high gains in classification accuracy and recall relative to the standard deep learning models. In the case of CIFAR-10, the hybrid model that used CNN to extract features and logistic regression to classify the features enhanced per-class and overall model robustness. The CNNs performed well in the extraction of spatial attributes in the images, whereas the logistic regression improved the capacity of the classifier to provide final predictions on the basis of the spatial attributes.

Equally, in UCI HAR, the CNN-LSTM hybrid model was highly accurate in the recognition of human activities and the performance of the CNN-LSTM was much higher compared to the DL-only approaches. The CNNs learned the spatial characteristics of the sensor data and the LSTMs learned the time-dependence of the sequential data, which

is a vital part in human activity recognition activity. This is in line with Kumar et al. [5] and Reyes-Ortiz et al. [6] who showed the usefulness of using CNNs in conjunction with LSTMs to process sequential data.

Image classification with CNNs is not new, and specifically with CIFAR-10, Goodfellow [14] LeCun et al. [15] demonstrated the power of the deep-learning approach in automatically learning image features. Nonetheless, they did not explore hybrid models that use CNNs with machine learning classifiers, which this study has revealed to be effective in both image classification and human activity recognition tasks.

The principle of hybrid models to human activity recognition is not recent because the previous researches have proven the viability of using CNNs and LSTMs together as a hybrid method to HAR and reported that the hybrid approach to HAR with the combination of CNNs to extract features and LSTMs to model the sequence improves the overall performance. It trained CNN-LSTM hybrids on HAR and achieved accuracy of activity recognition. Their findings are supported by this research, which indicates that hybrid models result in more precise predictions, particularly in dynamic settings where temporal relationships have a significant role.

Moreover, Sánchez Hernandez et al. [20] have examined the significance of the hybrid models in image classification by coming up with an ensemble architecture based on deep learning models to classify fine-grained images. The EDICA architecture is a combination of several deep learning models that can improve the performance of the CIFAR-10 classification, which is directly related to the hybrid approach that is used in the present study.

This study has a number of implications both to theory and practice. In theory, this study supports the idea that hybrid models, in particular, those based on the combination of CNNs to extract features and machine learning classifiers, are more efficient than the DL-only models in the activities that demand high classification accuracy and generalization. The CNNs are useful to capture the complicated characteristics in images and sensor data, whereas machine learning classifiers enhance the decision-making process, which makes this method very flexible to various data types and domains.

In practice, the hybrid models discussed in this paper can be used in a broad range in smart healthcare systems, autonomous vehicles, and surveillance systems, where both sensor and visual data must be processed effectively. As an illustration, these hybrid models could be used in smart healthcare systems to track patient behaviours and identify abnormal behaviours, which enhance patient care and shorten response time. Moreover, CNN-LSTM hybrid model can be used in the real-time activity recognition systems where the capability to process dynamic sensor data and correctly identify activities is essential.

Although these are encouraging findings, this study has a number of limitations. The computational complexity of training hybrid models, and especially those based on deep learning and machine learning strategies, is one of the primary constraints. The CNN-LSTM hybrids are very expensive in terms of computational resources particularly when they are being trained on large datasets such as CIFAR-10 and UCI HAR. This can restrict their usefulness in resource limited systems, e.g. edge devices or real time applications.

Also, the hybrid models worked well on the CIFAR-10 and UCI HAR datasets, but it is still not clear how they can be generalized to other datasets. These models have not been proven to work well on other image or time-series datasets, and they should be studied further.

The next step in the research should be to maximize the computational efficiency of the hybrid models in order to make them more applicable to real-time application. This can be done by the use of methods such as model pruning, quantization, and knowledge distillation which can shrink the size and complexity of the deep learning models without compromising the performance.

The other avenue that can be explored to work in future is the application of attention mechanism in CNN-LSTM hybrid models. Attention mechanisms enable the models to concentrate on the most important aspects of the input data that may enhance the accuracy of the activity recognition, particularly in the situation when the data includes irrelevant or noisy information. Also, the methods of transfer learning might be considered to take advantage of already trained models and enhance the generalization of hybrid models on new, unseen data. Lastly, more advanced systems could be developed with the implementation of multi-modal models that combine vision, sensor information, and even audio to approach more complex tasks in the real world, e.g. smart cities or autonomous vehicles.

6. Conclusion

The efficacy of hybrid deep learning and machine learning models in picture classification and human activity recognition (HAR) tasks is highlighted in this research. In comparison to standalone deep learning models, the results show that using CNNs to extract features and traditional machine learning classifiers, such as logistic regression, can significantly enhance a model's performance, leading to higher accuracy and improved recall. In particular, the CIFAR-10 and UCI HAR datasets were utilized to demonstrate the benefits of hybrid models in the image classification and sensor-based activity recognition, respectively. The implications of these findings on the real world are significant because they can be used in areas like healthcare, autonomous systems and surveillance where both visual and sensor data must be effectively processed. Nevertheless, there are still problems in getting the computational efficiency of these models to be optimized, particularly in real-time applications in resource-constrained environments. To advance the future research, it is recommended to enhance the generalization of

hybrid models on a variety of datasets, implement attention mechanisms to select features in a better way, and consider such techniques as transfer learning to improve the performance on unseen data. Also, in future research, multi-modal learning and edge computing may enhance hybrid models to be more applicable and applicable in the real world.

References

1. K. Bayoudh, "A survey of multimodal hybrid deep learning for computer vision: Architectures, applications, trends, and challenges," *Information Fusion*, vol. 105, p. 102217, 2024.
2. P. Kumar, S. Chauhan, and L. K. Awasthi, "Human activity recognition (HAR) using deep learning: Review, methodologies, progress and future research directions," *Archives of Computational Methods in Engineering*, vol. 31, no. 1, pp. 179–219, 2024.
3. S. Abbaspour, F. Fotouhi, A. Sedaghatbaf, H. Fotouhi, M. Vahabi, and M. Linden, "A comparative analysis of hybrid deep learning models for human activity recognition," *Sensors*, vol. 20, no. 19, p. 5707, 2020.
4. M. Ezzeldin, S. Ghoneim, A. Abdelhamid, and A. Atia, "Multi-modal hybrid hierarchical classification approach with transformers to enhance complex human activity recognition," *Signal, Image and Video Processing*, vol. 18, no. 12, pp. 9375–9385, 2024.
5. I. U. Khan, S. Afzal, and J. W. Lee, "Human activity recognition via hybrid deep learning based model," *Sensors*, vol. 22, no. 1, p. 323, 2022.
6. J. L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, 2016.
7. D. N. Maurya, D. Arora, and C. P. Singh, "A Hybrid CNN-LSTM Deep Neural Network Model for Efficient Human Activity Recognition," in *Int. Conf. Deep Learning and Visual AI*, Singapore: Springer Nature Singapore, 2024, pp. 403–411.
8. O. Alashram, N. Alagha, M. AlKakuri, Z. Swaveel, and A. Copiaco, "Hybrid Deep Feature Extraction and ML for Construction and Demolition Debris Classification," in *Proc. 8th Int. Conf. Signal Processing and Info. Security (ICSPIS)*, Nov. 2025, pp. 1–5.
9. D. Akinola, A. O. Oyedemi, and M. O. Ajinaja, "A deep learning-based hybrid CNN-LSTM model for human activity recognition," *Int. J. Sci. Res. in Computer Science and Engineering*, vol. 12, no. 6, 2024.
10. M. R. Atia, "ATTENTION-ENHANCED CNN-LSTM MODELS FOR SENSOR-BASED HUMAN ACTIVITY RECOGNITION: A COMPARATIVE STUDY," 2025.
11. A. Barua, M. U. Ahmed, and S. Begum, "A systematic literature review on multimodal machine learning: Applications, challenges, gaps and future directions," *IEEE Access*, vol. 11, pp. 14804–14831, 2023.
12. A. Younesi, M. Ansari, M. Fazli, A. Ejlali, M. Shafique, and J. Henkel, "A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends," *IEEE Access*, vol. 12, pp. 41180–41218, 2024.
13. Y. Abouelnaga, O. S. Ali, H. Rady, and M. Moustafa, "Cifar-10: KNN-based ensemble of classifiers," in *Proc. 2016 Int. Conf. Computational Science and Computational Intelligence (CSCI)*, 2016, pp. 1192–1195.
14. I. Goodfellow, *Deep Learning*, MIT Press, 2016.
15. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
16. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
17. C. Zhang, Z. Yang, X. He, and L. Deng, "Multimodal intelligence: Representation learning, information fusion, and applications," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 3, pp. 478–493, 2020.
18. J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
19. B. Jena, S. Saxena, G. K. Nayak, L. Saba, N. Sharma, and J. Suri, "Artificial intelligence-based hybrid deep learning models for image classification: The first narrative review," *Computers in Biology and Medicine*, vol. 137, p. 104803, 2021.
20. J. P. Sánchez Hernández, A. J. González Hernández, J. F. Solis, D. L. Hernández Rabadán, J. González-Barbosa, and G. Castilla Valdez, "EDICA: A Hybrid Ensemble Architecture Using Deep Learning Models for Fine-Grained Image Classification," *Mathematics*, vol. 13, no. 22, p. 3729, 2025.
21. P. Mavaie, L. Holder, and M. K. Skinner, "Hybrid deep learning approach to improve classification of low-volume high-dimensional data," *BMC Bioinformatics*, vol. 24, no. 1, p. 419, 2023.
22. A. Krizhevsky, V. Nair, and G. Hinton, "CIFAR-10 and CIFAR-100 datasets," Department of Computer Science, University of Toronto, 2009. [Online]. Available: <https://www.cs.toronto.edu/~kriz/cifar.html>
23. J. Reyes-Ortiz, D. Anguita, A. Ghio, L. Oneto, and X. Parra, "Human Activity Recognition Using Smartphones Dataset," UCI Machine Learning Repository, 2013. [Online]. Available: <https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones>