Deep learning approaches, platforms, datasets for behavior-based recognition: a survey

Yunusa Mohammed Jeddah[1], Aisha Hassan Abdalla Hashim[1,2], Othman Omran Khalifa[1,3]
[1]Department of Electrical and Computer Engineering, International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia
[2]Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, South Africa
[3]Libya Libyan Center for Engineering Research and Information Technology, Bani Walid, Libya

Article Info

Article history:

Received Aug 28, 2024
Revised Nov 28, 2024
Accepted Feb 27, 2025

Keywords:

Behavior recognition
Deep learning
Human activity recognition
Machine learning
Video surveillance

ABSTRACT

Video surveillance is an extensively used tool due to the high rate of atypical behavior and many cameras that enable video capture and storage. Unfortunately, most of these cameras are operator dependent for stored content analysis. This limitation necessitates the provision of an automatic behavior identification system. This behavior identification can be achieved using unsupervised (generative) computer vision methods. Deep learning makes it possible to model human behavior regardless of where they could be. We attempt to classify current research work to report the ongoing trends in human behavior recognition using deep learning algorithms. This paper reviews various aspects, like the ones associated with machine learning and deep learning models, human activity recognition (HAR), deep learning frameworks/tools, abnormal behavior datasets, and a variety of other current trends in the field of automatic learning. All these are to give the researcher a sense of direction in this area.

Corresponding Author:

Yunusa Mohammed Jeddah
Department of Electrical and Computer Engineering, International Islamic University Malaysia (IIUM)
Gombak, Kuala Lumpur, Selangor, Malaysia
Email: yunusmj2@hotmail.com

1. INTRODUCTION

In recent years, there has been an increase in research focused on AI (artificial intelligence), particularly the deep learning area. This surge in research interest has led to several studies surveys, and review papers which explore various deep learning applications. Notably, deep learning has been employed in the field of medicine such as eye disease diagnosis with the use of retinal fundus images [1] and for improving radiology techniques [2]. Other prominent research area involves human activity recognition (HAR), in which deep learning algorithms are increasingly employed in human behavior recognition, including detection of abnormal behavior [3], and distraction recognition [4]. Most recently, the trend has shifted towards anomalous behavior recognition, particularly in security and surveillance contexts [3].

Despite the advancements made, there are still many challenges and obstacles that remain. Existing surveys regularly focus on particular deep learning architectures or specific applications, such as HAR, without offering a broad overview of available approaches, platforms/tools, and datasets for behavior recognition. The diversity in data and behaviors involved further limits the ability of deep learning approaches to generalize over different situations and environments, despite their impressive performance in predicting and recognizing human behaviors.

The main aim of this survey is to present a comprehensive survey of deep learning-based behaviour recognition approaches, platforms/tools, and datasets. This survey attempts to assist researchers, especially new researchers in the field of behavior recognition by acquainting them with various deep learning approaches, platforms, and dataset that can potentially be adopted. Our study is intended as a comprehensive survey to give an overview of current approaches and trends in behavior recognition using deep learning. It does not follow the systematic and exhaustive methods of a systematic literature review (SLR) or a scoping review, but rather seeks to broaden the landscape, identify common themes and highlight key techniques. In contrast to previous surveys which are often limited in scope, our survey provided a broader perspective on the uses of deep learning security and surveillance, covering both the identification of suspicious or abnormal behaviors and the recognition of normal human behaviors.

Deep learning overcomes the limitations of machine learning by autonomously extracting the features from raw data and capturing complex temporal relationships in dynamical environments. These capabilities make deep learning especially very suitable for human behavior recognition, since it makes it possible to extract high-level representations from motion sensors and other source data. This research field has demonstrated rapid progress over the past decade, as demonstrated by a shift from traditional machine learning algorithms to deep learning approaches from 2006 to 2018 [5]. This trend highlights how researchers are increasingly adopting and accepting deep learning models. Figure 1 illustrates the shift in researchers' focus from traditional algorithms towards deep learning techniques. At the beginning, traditional methods were dominant, peaking up in year 2013. However, their predominance dropped as deep learning gained acceptance in the year 2014, going beyond traditional techniques by the year 2005 and continue to dominate onwards.

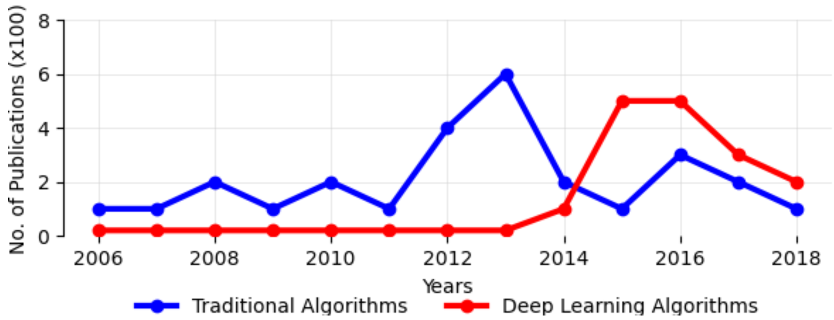

Figure 1. Traditional machine learning vs deep learning algorithms used per year [5]

The contributions of our survey include the following:

i) An in-depth review of deep learning approaches and platforms/tools.
ii) An elaborate description of the most popular datasets related to behavior recognition.
iii) Discussion on deep learning models and their specific use cases.
iv) A concise and clear guide that can serve as a starting point for new researchers in the field of deep learning for behavioral analysis.

To achieve this, we thoroughly conducted a general literature search to come up with a comprehensive survey of the current landscape in behavior recognition using deep learning. We used targeted key terms search terms linked to behavior recognition, deep learning approaches, platforms/tools, and datasets across several academic databases to harvest a wide range of related studies and resources. The downloaded papers were broadly checked and categorized by common subjects, like type of algorithm, application areas, and relevant dataset. The main focus of this survey is to provide an advanced level assessment and discussion of the various approaches, emphasizing important use cases and trends as opposed to sticking to rigid standards or comprehensive inclusion standards. Mendeley was employed to efficiently manage references from the identified sources and papers. The organization of the paper is as follows: in section 2, related work is presented. in section 3, human behavior recognition using deep learning techniques is discussed; in section 4, several deep learning platforms and tools are covered; section 5 presented datasets for behavior recognition; Results and discussion is presented in section 6, and in section 7, the paper is concluded.

2. RELATED WORK

Recent surveys have explored deep learning approaches for anomaly detection and human behavior recognition, especially in video surveillance contexts. Himeur *et al.* [6] presented a thorough review focusing on employing domain adaptation and deep transfer learning methods to make improvement in generalization in video surveillance systems. Their work is however, mostly limited to surveillance applications and does not explore other contexts. Similarly, Wastupranata *et al.* [7] focuses mainly on surveillance applications, stressing on methods for detecting abnormal behaviors but not offering a wider overview of deep learning models for diverse behavior recognition applications.

A number of other studies concentrate on surveillance videos anomaly detection using deep learning. Choudhry *et al.* [8] and Jebur *et al.* [9] largely highlight anomaly detection, mostly neglecting addressing the diversity of datasets, platforms, and use cases that are involved in general behavior recognition. A survey by [10] further strengthens this trend by concentrating narrowly on abnormal instances detection, failing to expand to other domains. In a more focused context, Shubber and Al-Ta'i [11] and Pham *et al.* [12] focus on certain behavior recognition undertakings such as human action recognition and violence detection. Despite providing valuable insights, these studies are restricted to certain activities and do not consider the full range of deep learning approaches and platforms. In addition, [13] focuses on healthcare tasks, primarily analyzing anomaly detection in daily actions but lack in covering of broader contexts like public security.

Contrary to these current studies, our own paper attempts addressing the gaps identified by providing a broad overview of deep learning approaches, platforms/tools, and datasets used for behavior recognition. Unlike previous reviews, which scope is often limited to specific applications, our study offers a wider perspective on the application of deep learning in behavior recognition, comprising several deep learning algorithms, various platforms, and a wide range of datasets. This all-inclusive approach will help in guiding researchers in choosing the right method and tool for different research circumstances and recognizes future research guidance not broadly covered in prior studies.

3. DEEP LEARNING APPROACHES TO HUMAN BEHAVIOUR RECOGNITION

Many research surveys have been conducted on human activity/behavior recognition through different facets, while other researchers focused on general assessments of human activity/behavior. The various aspects studied by the researchers include the following: the algorithm types, the methods used, the type of sensors used, the device type, and the type of activity performed. Here, we focus on a survey on the approaches to human behavior recognition. On a general note, there is an increase in researchers that try to apply deep learning approaches in studying human behavior. Human behavior is extraordinarily complex; human behavior is triggered by habits or intentions; transformed by effect, skill, and attitude; and affected by contextual and physical conditions [14]. Aguileta *et al.* [15] explored a range of techniques and approaches put forth by researchers to merge data from different sensors to discover research prospects in this area. Despite being in-depth and extensive, the survey needed to provide details about the implementation.

In their survey, [16] based their HAR review on traditional machine learning techniques. While [17] focused on a device-free method based on radio-frequency identification (RFID) technology. The device-free method does not require subjects to carry or wear a device for the sensors to recognize activities. Instead, sensors like RFID and cameras tag both the objects and the environment. This approach has advantages but is complicated in real-life implementation and has significant privacy-related issues. Shoaib *et al.* [18] categorized activities into two, complex activities and simple activities. Simple ones are easy to identify using a sensor because they are naturally repetitive, like standing, sitting, running, smoking, eating, and making coffee. While on the contrary, complex activities are difficult to recognize, as they are not recurrent and require more effort and data. Some datasets are intended for complex activities, like [19]; others are designed only for simple activities, like [20], while others are made for both activity types, such as [21].

Behavior recognition is a vital sub-area of HAR. The main idea is to identify a person's behavior from the data gotten via different sensors. Behavior detection is instrumental in several circumstances, like intelligent surroundings (aged care centers and smart homes) [19] and shopping centers. For example, in aged care centers, patients can be remotely monitored, significantly reducing the cost involved. By extension, automatic behavior recognition can be achieved through the use of deep learning and machine learning.

There are several machine learning-based models like k-nearest neighbor (KNN), suitable for classification problems [22], support vector machine (SVM), for handwriting recognize, identify a speaker, recognize fraudulent credit cards, as well as detect face [23], decision tree (DT) known for enhanced performance outcome's view [24] and many others have been used for HAR. However, using traditional algorithms has decreased greatly after the remarkable performance of deep learning-based algorithms in recognizing human action.

3.1. Deep learning methods

Several deep models, among which are convolutional neural network (CNN), deep neural network (DNN), recurrent neural network (RNN), restricted Boltzmann machine (RBM), long-short term memory

(LSTM), and many hybrids deep models (combining more than one deep structure), have been the subject of research studies [25]-[27]. Choice of the best suitable deep learning or machine learning model relies on the complexity and the problem type. The section looks at the most employed types of deep learning algorithms described in the following subsection.

3.1.1. Deep neural networks

DNN, as depicted in Figure 2 is an advanced type of artificial neural network (ANN) [28]. They are termed deep for having deep architecture by possessing lots of hidden layers sandwiched between the input and the output layers [29]. ANN having shallow networks has less hidden layers and is identified as multilayer perceptron (MLP). Researchers use DNN for human activity recognition and hand engineering features extracted from sensors. Similarly, [30] utilized principal component analysis (PCA) for feature extraction and DNN to learn the activities. Previous research has demonstrated that activities can be identified from raw sensor data when a structure is deep enough and has enough data, negating the need for custom characteristics [31]. A policy gradient-based reinforcement learning (RL) approach was utilized to accomplish online training for a DNN-based online activity identification system that was proposed by [32]. For the RL machine learning paradigm, learning is accomplished by rewards and punishments in a stochastic environment. RL has also employed deep learning to estimate the policy and reward function [33].

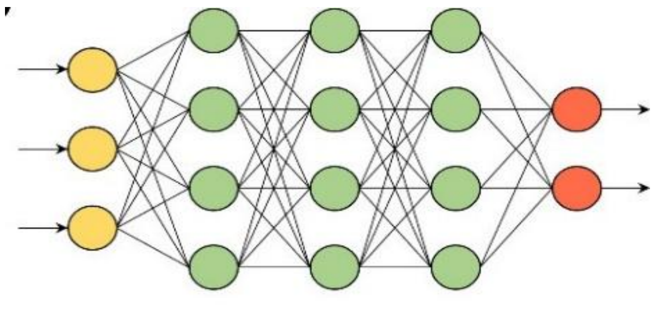

Figure 2. Description of DNN model [34]

3.1.2. Auto-encoder

As an unsupervised learning-based, auto-encoder (AE) [35], [36] is a neural network [37], [38] that uses the backpropagation algorithm. To distinguish between normal and abnormal network events, [39] proposed the use of AE and statistically analysis-driven intelligent intrusion detection system (IDS). Three layers make up AE's architecture: an input, a hidden (encoding) layer, and a decoding layer [40]. It was applied in many fields, like image classification, natural language processing, face recognition, and other areas [41], achieving impressive results. It was employed by [35] to propose a plug-and-play, “Kitsune”, that can identify potential network attacks without supervision. Figure 3 shows the AE’s structure, showing the processes of encoding and decoding.

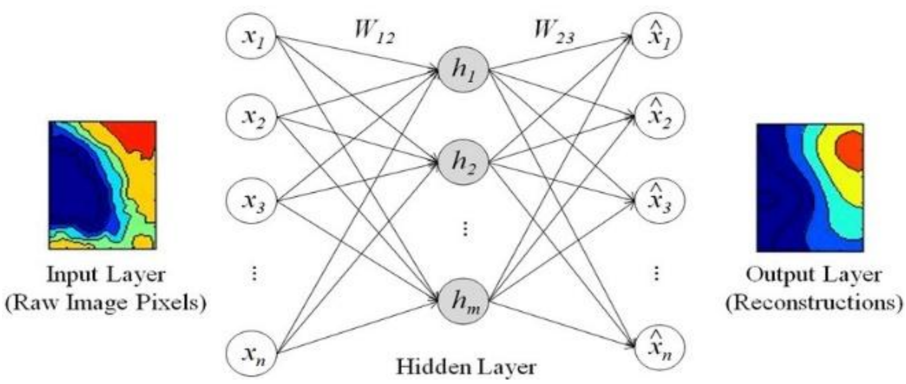

Figure 3. Structure of an AE [42]

3.1.3 Convolution neural networks

CNN models are mostly employed for their ability in feature learning from new sensor data [43]. When the input data has a clean spatial structure, CNN is a good choice (e.g., a collection of pixels in a photo). In addition, CNN has been used to recognize human activity in several study investigations [44]–[46]. Rueda *et al.* [44] asserted that numerous packed layers of convolution filters and pooling processes can be used to learn nonlinear and temporal structures in both simple and complex human movements. Ignatov [45] states, CNN was utilized for introducing a user-independent and online approach to recognizing human activities. Alongside feature extraction, the window size was also investigated, where it was discovered that a larger window size only improves performance in some activity classification cases.

3.1.4 Restricted Boltzmann machines

RBM (with its basic structure shown in Figure 4), proposed in 1986 by Paul Smolensky [47], trained to recreate its input from the first hidden layer in unsupervised pre-training. RBM is utilized in collaborative filtering [48], dimensionality reduction [49], classification [50], regression [49], feature learning [51], and topic modelling [47]. In their research, [37] employed RBM for pre-training layer by layer to obtain the initial weight and offset in their image reconstruction.

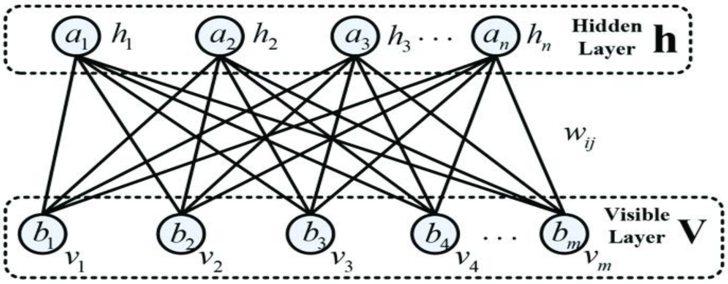

Figure 4. The basic structure of RBM [52]

3.1.5. Multilayer perceptron

MLP is a perceptron where groups with many perceptions are deployed layers wise to fix complex issues [53]. MLPs are three-layered [54], wherein the first layer is an input layer, that transmits signal to the hidden (second) layer, which also transmits the signal to the output layer. The primary use of MLP is the prediction of labels or classes-based classifications. A simple general idea of the structure of MLP is illustrated in Figure 5.

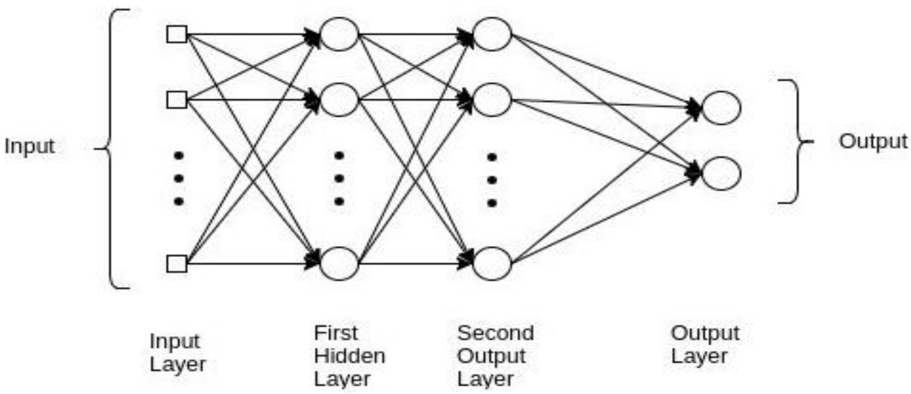

Figure 5. A simple overview of MLP [55]

3.1.6. Recurrent neural networks

Unlike conventional algorithms, RNN does not presume that the data sequences are different from each other as the prior sequences' information is utilized to learn the current sequence [56], as it could be seen in Figure 6. Therefore, RNN is frequently utilized to learn a time series' temporal structures and dynamics. Gated recurrent unit (GRUs) and LSTM are two most common variants of RNN where LSTM variant is seen as a better alternative, as it was formerly reported to give better results for problems related to HAR.

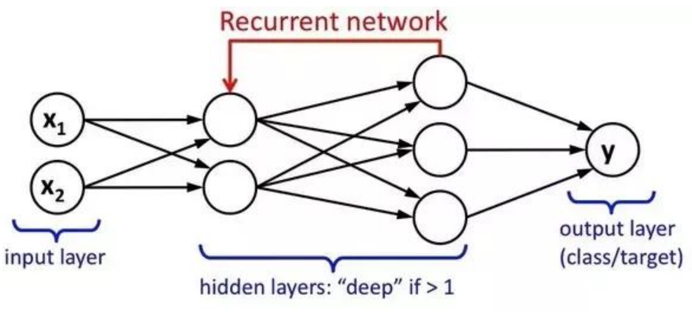

Figure 6. RNN structure [56]

A study by [57] investigated three types of RNN; LSTM, vanilla RNN (VRNN.), and GRU. Their effectiveness for activity recognition and detecting abnormal dementia-suffering patients' behavior was explored. While a deep stacked LSTM was proposed in [58] with the capability of learning and generalizing temporal dynamics from raw data. The structures of the GRU and LSTM, schemes can be seen in Figure 7. This diagram is crucial to comprehend the advances in RNN architectures, which have been made to improve the capacity to learn long-term relationships and to deal with issues such as vanishing gradients.

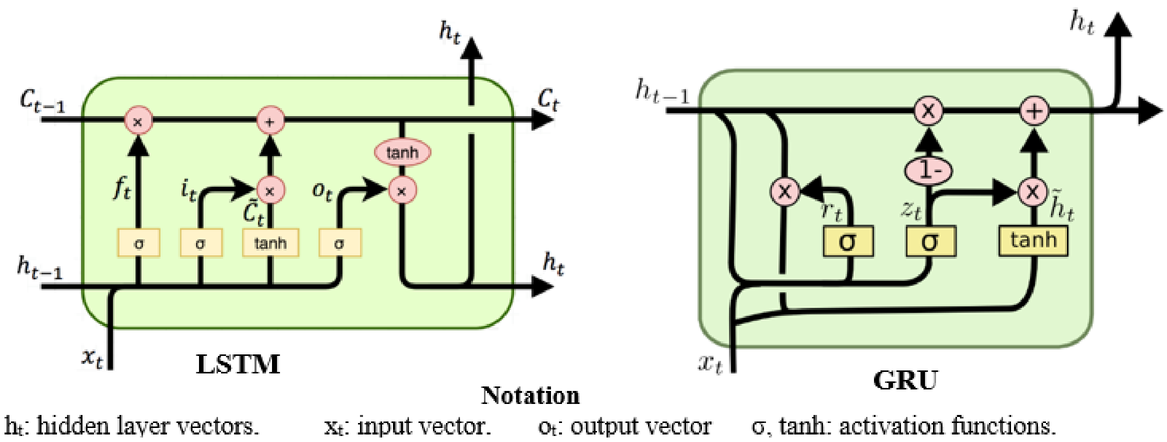

Figure 7. LSTM, GRU schemes [59]

3.1.7. Hybrid models

These are exceptional cases where two or more deep structures are combined. Combining two or more [60] deep learning models is to take advantage of the deep learning structures involved. For example, CNN, known for capabilities in feature extraction can be concatenated with the LSTM model, which can learn temporal dynamics to accomplish both features. The applications of hybrid models are not restricted to HAR. Still, they have applied in a variety of applications domains like visual recognition [61], voice (speech) [62], computer vision and natural language processing [63] and medical diagnosis [64].

An extensive summary of the main deep learning models and their application domains is given in Table 1. However, the choice between the deep learning architectures outlined in Table 1 depends on the use case at hand and the available data. In nutshell, CNN is regarded as the best choice when it comes to tasks related to images [65], RNN is ideal for sequential data like speech recognition, and natural language processing [66]. Others also have their respective strong use case strength.

Table 1. Deep learning models and their major application areas and description

Architecture	Major application areas	Description
CNN	Natural language processing, document analysis, image recognition, face recognition, video analysis	Utilizes convolutional layers to slide over the input information for feature learning.
AE	Understanding compact representation of data, natural language processing	A non-supervised neural network employed for learning data coding from samples.
RNN	Speech and handwriting recognition	Comprises of loops that provide information storage within the network.
RBM	feature learning, collaborative filtering, dimensionality reduction, classification, regression, and topic modelling	A generative stochastic NN that understands probability distribution across samples.
MLP	Image recognition, speech recognition, and machine translation	A feedforward artificial ANN

3.2. Deep learning object detection algorithms

Object detection is a precious and popular computer vision method that deals with object classification and localization in an image or video. Deep learning-based object detection approaches use CNN architectures such as you only look once (YOLO), single shot detector (SSD), and region proposals. Object detection is applied in asset inspection, pedestrian detection, self-driving cars, or video surveillance. This subsection explores the deep learning models employed for object (human in this case) detection. Furthermore, object detection can be employed in intelligent video analytics (IVA) wherever CCTV cameras are situated to study how people interact with one another and their surroundings.

3.2.1. YOLO

YOLOv1 utilizes a single neural network [67] as a real-time object detection system. According to [68], YOLO object detection is over one thousand times faster than region-based convolutional neural networks (R-CNN) and a hundred times faster than fast R-CNN. There exist different versions of YOLO: YOLOv2 [69], YOLOv3 [70], YOLOv4 [71], and YOLOv5 [72].

3.2.2. Single-shot detector

A renowned one-stage object detector [73], SSD can predict multiple classes faster than YOLO [74]. It is easy to train and incorporate with software systems requiring object detection module. SSD has superior accuracy compared to other single-stage techniques, even with smaller input image sizes [74].

3.2.3. Region-based convolutional neural networks

Established as a de facto deep learning-based object detection algorithm, R-CNN [75], [76] are revolutionary approaches which employ deep models to detection of object. The major weakness of R-CNN is that it is slow to implement. Fast R-CNN was developed in 2015 [77] with an aim to significantly reduce train time and drastically solve the limitations of R-CNN. However, YOLO is still a faster option due to the ease of the code [68].

3.2.4. SqueezeDet

Famous research within the scope of unified, low power, small CNN proposed SqueezeDet for object detection [78]. In the research, the authors show that SqueezeDet is a lightweight CNN explicitly developed for real time object detection, which utilizes computer vision techniques to detect objects. Similar to YOLO, SqueezeDet is a single-shot detector algorithm inspired by YOLO and SqueezeNet which creates and classifies the entrant region proposals in one neural network.

3.2.5. MobileNet

MobileNet [79] is also a SSD network that runs detection of object tasks. MobileNet is utilized by the Caffe framework. It is a lightweight deep CNN that is much smaller in size and performs much faster than many popular neural network models. Its main applications are the classification and detection of images. It utilizes depth-wise, separable convolutions, which involve performing a single convolution on each color channel instead of combining all three and flattening it.

3.2.6. Generative Adversarial Networks

Generative adversarial networks (GANs) were convincingly described as a powerful approach to unsupervised learning in machine learning, where two neural networks (the generator and the discriminator) compete in a game like manner in research conducted by [80]. The authors explained that the generator creates samples that are targeted at replicating the data distribution, whereas the discriminator aims to differentiate between true and generated data. This adversarial approach allows GANs to acquire complex data distributions, making them specifically effective for tasks like generating real images and transforming data across different realms. In their investigation, Posilović *et al.* [81] employed GAN to generate actual data, which can improve object detectors.

3.2.7. You only learn one representation

You only learn one representation (YOLOR) [82] is a novel algorithm for detection of object, different to the YOLO due to differences of inventors, model infrastructure, and architecture. It aims to provide ability to machine learning algorithms to serve many other tasks given a single input [83]. While CNNs learn to analyze inputs to obtain the outputs, YOLOR attempts to have CNN's both (i) learn getting outputs and, at the same time and (i) know what all different outputs could be. Instead of only a single output, YOLOR can have a lot.

4. DEEP LEARNING TOOLS

Deep learning tools enable data researchers to build programs that may give a machine the ability to learn as a human brain and process data and patterns before making decisions. The tools rely on predictive modelling and statistics, which help data scientists gather, interpret, and then analyze enormous amounts of data. These tools can help to seamlessly detect objects, recognize speech, translate languages, and make decisions appropriately. Also referred to as deep learning platforms or frameworks, software packages were available to the researchers to mitigate deep learning architectures construction. However, some years ago, non-deep learning experts encountered many challenges in managing software packages. Such a circumstance continued until Google in 2012 proposed the DistBelief system. Next to that, software packages such as TensorFlow, DeepLearning4j, Microsoft Cognitive Toolkit (CNTK), Torch, Caffe, H2O.ai, Neural Designer, Keras, and Deep Learning Kit, have widely fueled the industry.

4.1. Neural designer

A user-friendly application developed by Artelnics, which allows building DNNs without coding or building block diagrams [84]. Amongst its applications, you can also find algorithms for different training strategies, data statistics and preparation, testing and deployment of the model or export the results to other tools like R or Python. Furthermore, neural designer is compatible with the most familiar databases and data files and can run the most extensive datasets. Moreover, you can import data and use the machine learning platform to enhance productivity. The primary objective of neural designer is to facilitate innovative organizations to employ artificial intelligence (AI), with a focus on its applications and not based on mathematics or programming. Furthermore, neural designer supports a whole modelling cycle, from preparing data to model production [85].

4.2. H2O.ai

Developed employing Java as core technology, H2O provides enormous flexibility to researchers. Thanks to H2O, anybody can easily apply predictive, machine learning, analytics, and deep learning to unravel complex problems [86]. It utilizes the most accustomed interface [87], an open-source framework with an easy-to-use web-based graphical user interface (GUI). This tool helps in real-time data scoring, and it is highly scalable.

4.3. DeepLearningKit

Apple employs DeepLearningKit framework on most of their products such as OS X, iOS, and tvOS [88]. It supports the pre-trained models on Apple's devices with GPUs for conventional OS X computers [89]. In addition, it uses deep CNN, such as image recognition. It is currently trained with the caffe deep learning framework [90], but the long-term objective is to support the use of other models such as Torch and TensorFlow.

4.4. Microsoft Cognitive Toolkit

This toolkit is commercially used to train deep learning systems to learn accurately as the human brain. An open-source which offers excellent scaling abilities, enterprise-level quality, accuracy, and speed. Microsoft products like Cortana, Skype, and Xbox Bing, use it for AI industry-level generation [90].

4.5. Keras

A Python-based application program interface (API) employed to run TensorFlow [91], Theano, and CNTK. Keras is an enhanced level API of TensorFlow utilized for model building and proves similar in stacking layers. On the other hand, it has minimal functionality. It was designed to enable quick experiments [92] and works with TensorFlow and Theano. The main advantage is that fast results can be obtained with just one concept.

4.6. TensorFlow

This is the most prominent Google libraries for deep learning and machine learning applications [93]. It is built to run on GPUs, CPUs, and internet of things (IoT) processors such as neural network stick, and Jetson Nano. Released in 2015 by Google but the stable version has been available since 2017 under the Apache Open-Source License [94].

4.7. ConvNetJS

ConvNetJS enables users to design and resolve neural networks by using JavaScript for neural networks (deep learning models) training entirely in browser [95]. It is an experimental reinforcement learning module built on deep Q learning, not needing additional software, installations, GPUs or compilers. For processing images, ConvNetJS can define and train convolutional networks.

4.8. Torch

With a GPU-based MATLAB-like environment, the torch [94] is an extremely effective open-source computing platform that supports deep learning techniques [94]. It is a high-powered N-dimensional array that features many routines for indexing, transposing, and slicing [96]. It has outstanding support for GPU and is embeddable to collaborate with Android and iOS.

4.9. DeepLearning4j

DeepLearning4j [94] is an open-source, free java-based deep learning library which provides an extensive solution to deep learning in different applications such as knowledge discovery and deep predictive mining on CPU’s and GPU’s graphics processing. It integrates the AI algorithms and methods appropriate for business intelligence, cyber forensics, predictive analysis, face recognition, network intrusion detection and prevention, anomaly detection, recommender systems, and lots more. Models from advanced deep-learning frameworks like Caffe, Theano, Keras, and TensorFlow can be imported [94].

4.10. MATLAB deep learning toolbox

MATLAB is a proprietary multi-paradigm mathematical computing environment, developed by MathWorks that allows manipulations of matrix, plotting functions and data, implementation of algorithms, creation of user interfaces, as well as interfacing with programs built using other languages. MATLAB utilizes toolboxes, especially deep-learning toolboxes to run programs [84]. In 2019, it was used to build a GAN, attention networks, and variational autoencoders. CNN, LSTM, and networks with 3D CNN layers can all be combined using the MATLAB deep learning toolkit [97].

A platform for creating and implementing DNNs with algorithms, pre-trained models, and APPs is provided by the MATLAB deep learning toolbox. LSTM networks for time series analysis and transfer learning have been a part of the MATLAB toolbox since 2017 [97]. Amongst the key plus of MATLAB is that there are multiple GPUs, cluster computing, parallel and cloud computing, for speeding up the processes of deep learning. Table 2 shows a summary of the deep learning tools discussed. Prominent deep learning tools are presented in Table 2, in tandem with details about their license, primary programming languages, release dates, and official websites. The table intends to provide researchers with immediate guidance as they choose the right tools to use in their behavior-based recognition research.

Table 2. Deep learning tools summary table

Framework	License	Core language	Release year	Homepage
Neural Designer	Proprietary software	C++	2014	https://www.neuraldesigner.com/
H2O	Apache2.0	R and Python	2015	https://h2o.ai/
DeepLearningKit	Apache2.0	C++	2015	http://deeplearningkit.org/
Microsoft Cognitive Toolkit	MIT	C++	2016	https://learn.microsoft.com/en-us/cognitive-toolkit/
Keras	MIT	Python	2015	https://keras.io/
TensorFlow	Apache2.0	C++ and Python	2015	https://www.tensorflow.org/
ConvNetJS	MIT	JavaScript	2014	https://cs.stanford.edu/people/karpathy/convnetjs/
Torch	BSD	C and Lua	2002	http://torch.ch/
DeepLearning4j	Apache2.0	Java	2014	https://deeplearni ng4j.org/
MATLAB Toolbox	Proprietary software	C, C++, Java, MATLAB	1992	https://matlab.mathworks.com/

5. BEHAVIOR RECOGNITION DATASETS

Large-scale database existence is a prerequisite for the efficient working of deep learning system. Here, we look at some datasets used for behavior recognition in deep learning. Unfortunately, most early deep-learning state-of-the-art works were trained with private large-scale databases. A situation like this has caused researchers to need more public datasets to replicate the findings of these works or make a comparison of their models.

5.1. UCF-Crime dataset

Specifically created to assist with recognizing anomalous activities in difficult environments, the UCF-Crimes dataset is invaluable for real-world anomaly identification in surveillance videos. According to [98], this dataset allows for the creation and testing of effective detection algorithms because it comprises a wide range of crime-related events that were recorded from various camera angles and settings. The UCF-Crimes dataset contributes to the development of anomaly detection techniques by offering comprehensive annotations and a wide range of scenarios, to enhance safety and security through more efficient surveillance systems. UCF-Crime dataset comprises of long uncut surveillance videos that contain thirteen real-life abnormalities, comprising assault, abuse, arrest, shoplifting, road accident, arson, fighting, burglary, shooting, explosion, stealing, robbery, and vandalism. These abnormalities were chosen because they have a substantial effect on the safety of the public.

5.2. HBD21 dataset- human behavior dataset 2021

Jayaswal and Dixit [99] is comprised of normal and abnormal action video files prepared with natural and artificial light conditions to ease recognition. The dataset is made up of 456 annotated videos. This dataset comprises four categories a) gun violence, b) sabotage violence, c) assault violence, and d) normal actions. Every category contained over one hundred video files. Each of the videos is 9 seconds long, and it is observed that such a length is enough to represent real action. The ratio of training to testing video files is 7:3. In other words, 70% training and 30% testing videos of the original dataset. In this enhanced version, the brightness of each video is increased at an appropriate threshold, which helps achieve better performance in the training and testing stages.

5.3. MoLa InCar AR: dataset for action recognition

MoLa InCar AR [100] dataset is used in training human action recognition in a vehicle focused on violence detection. Thermal, RGB, depth, and event-based data were all recorded for this dataset with a resulting 6, 400 sample videos and over three million frames, gathered from different sixteen subjects. The dataset includes fifty-eight action classes, which include neutral (non-violent) and violent activities.

5.4. KU-HAR

The KU-HAR dataset provides an extensive set of data for the development and testing of recognition algorithms, which is a major contribution to the field of human activity recognition. In research conducted by [101], they posed that the KU-HAR dataset is an open dataset containing eighteen different activities generated from ninety participants (seventy-five male and fifteen female) with the help of smartphone sensors (gyroscope and accelerometer). It contains a wide range of activities that were accurately captured and documented numerous aspects of human behavior and interactions. The dataset consists of 20,750 extracted subsamples and 1,945 unprocessed activities gathered directly from the subjects. The KU-HAR dataset contributes to the development of activity recognition systems by offering large and comprehensive samples, which are intended to improve the precision and practicality of models in various contexts.

5.5. Multi-gait and single gait datasets

Previous research in gait by [102] for recognition using the available dataset focused on single subjects. But in a real-time situation (like airports, subway stations etc.) a more significant number of persons walking in groups and occlusion factors affect gait recognition performance. Thus, this new dataset which focused on dynamic occlusion scenarios, was presented. They built two distinct categories of gait dataset, i.e., subjects walk in groups (SMVDU-multi-gait), and subjects move individually (SMVDU-single-gait). The major intent of this dataset is to examine the differences in gait patterns by the time the same subject walks as a group or on an individual basis and also recognize the target subject in multi-gait.

6. RESULTS AND DISCUSSION

The survey reveals that CNNs and RNNs deep learning algorithms are the most often employed for tasks such as human behavior recognition due to their strengths in handling image data and sequential patterns respectively. Figure 8 illustrates that CNN-based algorithms dominates in the field of research applications due to their robust feature extraction abilities, most especially in visual data applications. The RNN-based algorithms, particularly LSTM, are the next popular choice owing to their effective sequential data handling and temporal dependencies learning. Hybrid deep learning models, where strengths of two or more architectures are combined also gained attention which indicates the growing interest in more adaptive and versatile approaches.

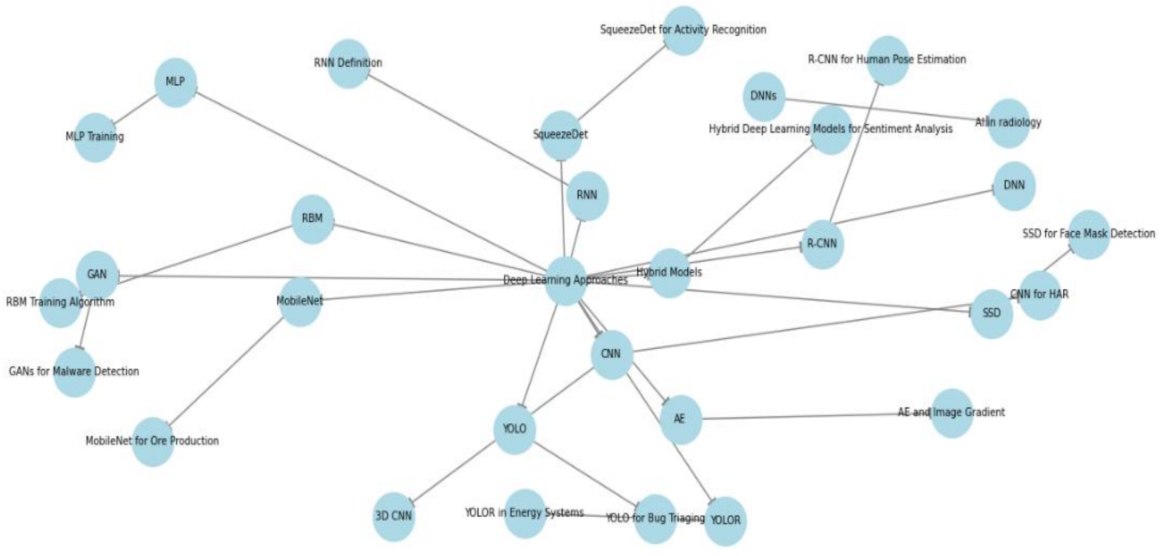

Figure 8. Summary of the major themes covered in this survey

In comparison to traditional algorithms, the survey findings are in alliance with existing literature which supports the greater deep learning algorithms performance in behavior recognition, most especially in complex activities and dynamic environments. That notwithstanding, whereas deep learning models offer remarkable advantages, they also face challenges such as the need for large, labelled datasets and high computational resources. An ongoing struggle for the development of more universal datasets for behavior recognition is indicated by the unexpected reality that context-specific augmentation requirements and inadequate labelling remain the main limitations of most datasets, which include UCF-Crime and KU-HAR.

This survey uncovers the multiplicity of deep learning approaches, platforms/tools, and datasets that are currently available for behavior recognition. The significance of this survey lies in its broad revelation of existing deep learning approaches and tools, offering a starting point for new researchers to choose appropriate approaches that are specific to their tasks and available datasets. Imminent research can explore the development of more universal datasets, optimizing hybrid architecture, and enhancing computational efficiency to advance the field further.

7. CONCLUSION

In this survey, we observed that unsupervised learning approaches are very suitable for human behavior recognition, primarily given the prevalence of unlabeled datasets. Nevertheless, relying on one approach is not recommended, as different algorithms can yield different results depending on the dataset used. Recently, hybrid models have been used to enhance either feature extraction or learning process, highlighting the need for tailored approaches for specific tasks. Researchers are recommended to consider performance metrics, like recognition accuracy. Due to the time-consuming nature of training on large video datasets, it is recommended to compress videos prior to feature extraction.

The study finds out that MATLAB's deep learning toolbox is a versatile tool, providing benefits like platform integration, access to the most recent research with the use of ONNX import features, as well as other pre-trained libraries such as NASNet, SqueezeNet, ResNet-101, and Inception-v3. The versatility of this tool is further enhanced by the ability to interoperate between MATLAB and Python.

Conclusively, this survey offers an extensive synopsis of deep learning approaches, platforms/tools, and datasets for behavior recognition; nonetheless, it is apparent that no platform or method is universally effective. When adopting an approach, researchers ought to carefully evaluate the specifics of their tasks, datasets, and anticipated outcomes. Leveraging several models, such as MACNN and RNN, contributes to improving the performance of the resulting algorithm.

ACKNOWLEDGEMENTS

This work is supported by the Ministry of Higher Education (MOHE) Fundamental Research Grant Scheme (FRGS22-264-0873) (Grant No: FRGS/1/2022/ICT11/UIAM/01/1).

FUNDING INFORMATION

This work was supported as mentioned in the acknowledgement section.

AUTHOR CONTRIBUTIONS STATEMENT

Yunusa M.J.'s contributions include Conceptualization, Methodology, Investigation, Data Curation, Writing (Original Draft and Review & Editing), Software, Validation, Formal Analysis, and Visualization. While Aisha H. A. H. contributed to the Methodology, Investigation, Resources, Validation, Formal Analysis, Writing (Review & Editing), Supervision, Project Administration, and Funding Acquisition. Othman O. K.'s contributions are Conceptualization, Methodology, Resources, Validation, Formal Analysis, Visualization, Supervision, and Project Administration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Yunusa Mohammed Jeddah	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Aisha Hassan Abdallah Hashim		✓		✓	✓	✓	✓			✓		✓	✓	✓
Othman Omran Khalifa	✓	✓		✓	✓		✓			✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Authors state no conflict of interest.

DATA AVAILABILITY

The data used in this review paper is available as duly referenced.

REFERENCES

[1] M. E. Hoque *et al.*, "A deep learning approach for retinal image feature extraction," *Pertanika Journal of Science and Technology*, vol. 29, no. 4, pp. 2543–2563, Oct. 2021, doi: 10.47836/PJST.29.4.17.
[2] E. Pakdemirli, "Artificial intelligence in radiology: friend or foe? Where are we now and where are we heading?," *Acta Radiologica Open*, vol. 8, no. 2, p. 205846011983022, Feb. 2019, doi: 10.1177/2058460119830222.
[3] A. Mehmood, "Abnormal behavior detection in uncrowded videos with two-stream 3D convolutional neural networks," *Applied Sciences (Switzerland),* vol. 11, no. 8, p. 3523, Apr. 2021, doi: 10.3390/app11083523.
[4] S. J. Jiao, L. Y. Liu, and Q. Liu, "A hybrid deep learning model for recognizing actions of distracted drivers," *Sensors*, vol. 21, no. 21, p. 7424, Nov. 2021, doi: 10.3390/s21217424.
[5] S. O. Slim, A. Atia, M. M. A. Elfattah, and M. S. M. Mostafa, "Survey on human activity recognition based on acceleration data," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 3, pp. 84–98, 2019, doi: 10.14569/IJACSA.2019.0100311.
[6] Y. Himeur *et al.*, "Video surveillance using deep transfer learning and deep domain adaptation: towards better generalization," *Engineering Applications of Artificial Intelligence*, vol. 119, p. 105698, Mar. 2023, doi: 10.1016/j.engappai.2022.105698.
[7] L. M. Wastupranata, S. G. Kong, and L. Wang, "Deep learning for abnormal human behavior detection in surveillance videos—a survey," *Electronics (Switzerland)*, vol. 13, no. 13, p. 2579, Jun. 2024, doi: 10.3390/electronics13132579.
[8] N. Choudhry, J. Abawajy, S. Huda, and I. Rao, "A comprehensive survey of machine learning methods for surveillance videos anomaly detection," *IEEE Access*, vol. 11, pp. 114680–114713, 2023, doi: 10.1109/ACCESS.2023.3321800.

[9] S. A. Jebur, K. A. Hussein, H. K. Hoomod, L. Alzubaidi, and J. Santamaría, "Review on deep learning approaches for anomaly event detection in video surveillance," *Electronics (Switzerland)*, vol. 12, no. 1, p. 29, Dec. 2023, doi: 10.3390/electronics12010029.

[10] H. T. Duong, V. T. Le, and V. T. Hoang, "Deep learning-based anomaly detection in video surveillance: a survey," *Sensors*, vol. 23, no. 11, 2023, doi: 10.3390/s23115024.

[11] M. S. M. Shubber and Z. T. M. Al-Ta'i, "A review on video violence detection approaches," *International Journal of Nonlinear Analysis and Applications*, vol. 13, no. 2, pp. 1117–1130, 2022, doi: 10.22075/ijnaa.2022.6369.

[12] H. H. Pham, L. Khoudour, A. Crouzil, P. Zegers, and S. A. Velastin, "Video-based human action recognition using deep learning: a review," *arXiv preprint 2208.03775*, Aug. 2022, [Online]. Available: http://arxiv.org/abs/2208.03775.

[13] N. C. Tay, T. Connie, T. S. Ong, A. B. J. Teoh, and P. S. Teh, "A review of abnormal behavior detection in activities of daily living," *IEEE Access*, vol. 11, pp. 5069–5088, 2023, doi: 10.1109/ACCESS.2023.3234974.

[14] A. A. Salah and B. Lepri, *Human behavior unterstanding*, vol. 7065. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[15] A. A. Aguileta, R. F. Brena, O. Mayora, E. Molino-Minero-re, and L. A. Trejo, "Multi-sensor fusion for activity recognition—a survey," *Sensors (Switzerland)*, vol. 19, no. 17, p. 3808, Sep. 2019, doi: 10.3390/s19173808.

[16] Z. Hussain, Q. Z. Sheng, and W. E. Zhang, "Different approaches for human activity recognition: a survey," *arXiv preprint 1906.05074*, pp. 1–28, Oct. 2019.

[17] W. S. Lima, E. Souto, K. El-Khatib, R. Jalali, and J. Gama, "Human activity recognition using inertial sensors in a smartphone: An overview," *Sensors (Switzerland)*, vol. 19, no. 14, p. 3213, Jul. 2019, doi: 10.3390/s19143213.

[18] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Complex human activity recognition using smartphone and wrist-worn motion sensors," *Sensors (Switzerland)*, vol. 16, no. 4, p. 426, Mar. 2016, doi: 10.3390/s16040426.

[19] W. Qi, H. Su, C. Yang, G. Ferrigno, E. De Momi, and A. Aliverti, "A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone," *Sensors (Switzerland)*, vol. 19, no. 17, p. 3731, Aug. 2019, doi: 10.3390/s19173731.

[20] A. Stisen *et al.*, "Smart devices are different: assessing and mitigating mobile sensing heterogeneities for activity recognition," in *SenSys 2015 - Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, Nov. 2015, pp. 127–140, doi: 10.1145/2809695.2809718.

[21] A. Toshev and C. Szegedy, "DeepPose: human pose estimation via deep neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1653–1660, doi: 10.1109/CVPR.2014.214.

[22] P. Cunningham and S. J. Delany, "K-nearest neighbour classifiers-a tutorial," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–25, Jul. 2021, doi: 10.1145/3459665.

[23] S. Huang, C. A. I. Nianguang, P. Penzuti Pacheco, S. Narandes, Y. Wang, and X. U. Wayne, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics and Proteomics*, vol. 15, no. 1, pp. 41–51, Jan. 2018, doi: 10.21873/cgp.20063.

[24] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[25] E. Acar, E. Şahin, and İ. Yılmaz, "Improving effectiveness of different deep learning-based models for detecting COVID-19 from computed tomography (CT) images," *Neural Computing and Applications*, vol. 33, no. 24, pp. 17589–17609, Jul. 2021, doi: 10.1007/s00521-021-06344-5.

[26] O. Gedik and A. Demirhan, "Comparison of the effectiveness of deep learning methods for face mask detection," *Traitement du Signal*, vol. 38, no. 4, pp. 947–953, Aug. 2021, doi: 10.18280/ts.380404.

[27] S. Mani, A. Sankaran, and R. Aralikatte, "Deeptriage: exploring the effectiveness of deep learning for bug triaging," in *ACM International Conference Proceeding Series*, Jan. 2019, pp. 171–179, doi: 10.1145/3297001.3297023.

[28] H. Guo, J. Zhou, M. Koopialipoor, D. J. Armaghani, and M. M. Tahir, "Deep neural network and whale optimization algorithm to assess flyrock induced by blasting," *Engineering with Computers*, vol. 37, no. 1, 2021, doi: 10.1007/s00366-019-00816-y.

[29] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: a survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, Mar. 2019, doi: 10.1016/j.patrec.2018.02.010.

[30] K. H. Walse, R. V. Dharaskar, and V. M. Thakare, "PCA based optimal ANN classifiers for human activity recognition using mobile sensors data," in *Smart Innovation, Systems and Technologies*, vol. 50, Springer International Publishing, 2016, pp. 429–436.

[31] Y. Bengio, "Deep learning of representations: looking forward," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7978 LNAI, 2013, pp. 1–37.

[32] G. Bhat, R. Deb, V. V. Chaurasia, H. Shill, and U. Y. Ogras, "Online human activity recognition using low-power wearable devices," in *IEEE/ACM International Conference on Computer-Aided Design, Digest of Technical Papers, ICCAD*, Nov. 2018, pp. 1–8, doi: 10.1145/3240765.3240833.

[33] T. Yang, L. Zhao, W. Li, and A. Y. Zomaya, "Reinforcement learning in sustainable energy and electric systems: a survey," *Annual Reviews in Control*, vol. 49, pp. 145–163, 2020, doi: 10.1016/j.arcontrol.2020.03.001.

[34] J. Baek and Y. Choi, "Deep neural network for predicting ore production by truck-haulage systems in open-pit mines," *Applied Sciences (Switzerland)*, vol. 10, no. 5, p. 1657, Mar. 2020, doi: 10.3390/app10051657.

[35] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: an ensemble of autoencoders for online network intrusion detection," *arXiv preprint arXiv:1802.09089*, 2018, doi: 10.14722/ndss.2018.23204.

[36] X. K. Li, W. Chen, Q. Zhang, and L. Wu, "Building auto-encoder intrusion detection system based on random forest feature selection," *Computers and Security*, vol. 95, p. 101851, Aug. 2020, doi: 10.1016/j.cose.2020.101851.

[37] X. J. Wu, M. Da Xu, C. Di Li, C. Ju, Q. Zhao, and S. X. Liu, "Research on image reconstruction algorithms based on autoencoder neural network of restricted Boltzmann machine (RBM)," *Flow Measurement and Instrumentation*, vol. 80, p. 102009, Aug. 2021, doi: 10.1016/j.flowmeasinst.2021.102009.

[38] H. He and J. Pathak, "An unsupervised learning approach to solving heat equations on chip based on Auto Encoder and Image Gradient," *arXiv preprint arXiv:2007.09684*, Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.09684.

[39] C. Ieracitano, A. Adeel, F. C. Morabito, and A. Hussain, "A novel statistical analysis and autoencoder driven intelligent intrusion detection approach," *Neurocomputing*, vol. 387, pp. 51–62, Apr. 2020, doi: 10.1016/j.neucom.2019.11.016.

[40] S. U. Jan, Y. D. Lee, and I. S. Koo, "A distributed sensor-fault detection and diagnosis framework using machine learning," *Information Sciences*, vol. 547, pp. 777–796, Feb. 2021, doi: 10.1016/j.ins.2020.08.068.

[41] A. Ashfahani, M. Pratama, E. Lughofer, and Y. S. Ong, "DEVDAN: deep evolving denoising autoencoder," *Neurocomputing*, vol. 390, pp. 297–314, May 2020, doi: 10.1016/j.neucom.2019.07.106.

[42] Y. Tao, X. Gao, A. Ihler, S. Sorooshian, and K. Hsu, "Precipitation identification with bispectral satellite information using deep learning approaches," *Journal of Hydrometeorology*, vol. 18, no. 5, pp. 1271–1283, Apr. 2017, doi: 10.1175/JHM-D-16-0176.1.

[43] F. Cruciani *et al.*, “Feature learning for human activity recognition using convolutional neural networks: a case study for inertial measurement unit and audio data,” *CCF Transactions on Pervasive Computing and Interaction*, vol. 2, no. 1, pp. 18–32, Jan. 2020, doi: 10.1007/s42486-020-00026-2.

[44] F. M. Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst, and M. T. Hompel, “Convolutional neural networks for human activity recognition using body-worn sensors,” *Informatics*, vol. 5, no. 2, p. 26, May 2018, doi: 10.3390/informatics5020026.

[45] A. Ignatov, “Real-time human activity recognition from accelerometer data using convolutional neural networks,” *Applied Soft Computing Journal*, vol. 62, pp. 915–922, 2018, doi: 10.1016/j.asoc.2017.09.027.

[46] G. de Aquino e Aquino, M. K. Serrão, M. G. F. Costa, and C. F. F. Costa-Filho, “Human activity recognition from accelerometer data with convolutional neural networks,” in *IFMBE Proceedings*, vol. 83, Springer International Publishing, 2022, pp. 1603–1610.

[47] Q. Wang, X. Gao, K. Wan, F. Li, and Z. Hu, “A novel restricted Boltzmann machine training algorithm with fast gibbs sampling policy,” *Mathematical Problems in Engineering*, vol. 2020, pp. 1–19, Mar. 2020, doi: 10.1155/2020/4206457.

[48] Z. Li and G. Li, “Improved collaborative filtering algorithm of restricted boltzmann machine,” in *Proceedings of 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2019*, Dec. 2019, pp. 328–334, doi: 10.1109/IAEAC47372.2019.8997592.

[49] J. Vrábel, P. Pořízka, and J. Kaiser, “Restricted Boltzmann machine method for dimensionality reduction of large spectroscopic data,” *Spectrochimica Acta - Part B Atomic Spectroscopy*, vol. 167, p. 105849, May 2020, doi: 10.1016/j.sab.2020.105849.

[50] C. Dewi, R. C. Chen, Hendry, and H. Te Hung, “Comparative analysis of restricted Boltzmann machine models for image classification,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12034 LNAI, Springer International Publishing, 2020, pp. 285–296.

[51] Z. Liu, *et al.*, “Research on unsupervised feature learning for Android malware detection based on restricted Boltzmann machines,” *Future Generation Computer Systems*, vol. 120, pp. 91–108, Jul. 2021, doi: 10.1016/j.future.2021.02.015.

[52] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, and Y. Zhao, “A decoding scheme for incomplete motor imagery EEG with deep belief network,” *Frontiers in Neuroscience*, vol. 12, Sep. 2018, doi: 10.3389/fnins.2018.00680.

[53] B. Turkoglu and E. Kaya, “Training multi-layer perceptron with artificial algae algorithm,” *Engineering Science and Technology, an International Journal*, vol. 23, no. 6, pp. 1342–1350, Dec. 2020, doi: 10.1016/j.jestch.2020.07.001.

[54] I. Lorencin, N. Anđelić, J. Španjol, and Z. Car, “Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis,” *Artificial Intelligence in Medicine*, vol. 102, p. 101746, Jan. 2020, doi: 10.1016/j.artmed.2019.101746.

[55] V. Arvind, “Detecting Emotions from Speech using MLP(Multilevel Perceptron) Algorithm,” Medium. [Online]. Available: https://medium.com/@vibhuarvind81/detecting-emotions-from-speech-using-mlp-multilevel-perceptron-algorithm-1b084bcc0892 (Accessed: Mar. 13, 2025)

[56] “Recurrent neural network,” *DeepAI*. https://deepai.org/machine-learning-glossary-and-terms/recurrent-neural-network (accessed Nov. 03, 2022).

[57] D. Arifoglu and A. Bouchachia, “Activity recognition and abnormal behaviour detection with recurrent neural networks,” *Procedia Computer Science*, vol. 110, pp. 86–93, 2017, doi: 10.1016/j.procs.2017.06.121.

[58] J. Donahue *et al.*, “Long-term recurrent convolutional networks for visual recognition and description,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 677–691, Apr. 2017, doi: 10.1109/TPAMI.2016.2599174.

[59] R. Saidi, F. Jarray, and M. Alsuhaibani, “Comparative Analysis of Recurrent Neural Network Architectures for Arabic Word Sense Disambiguation,” International Conference on Web Information Systems and Technologies, WEBIST - Proceedings, 2022, pp. 272–277, doi: 10.5220/0011527600003318.

[60] C. N. Dang, M. N. Moreno-García, and F. De La Prieta, “Hybrid deep learning models for sentiment analysis,” *Complexity*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/9986920.

[61] T. H. S. Li, P. H. Kuo, T. N. Tsai, and P. C. Luan, “CNN and LSTM based facial expression analysis model for a humanoid robot,” *IEEE Access*, vol. 7, pp. 93998–94011, 2019, doi: 10.1109/ACCESS.2019.2928364.

[62] V. Passricha and R. K. Aggarwal, “A hybrid of deep CNN and bidirectional LSTM for automatic speech recognition,” *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1261–1274, Mar. 2020, doi: 10.1515/jisys-2018-0372.

[63] R. Yang *et al.*, “CNN-LSTM deep learning architecture for computer vision-based modal frequency detection,” *Mechanical Systems and Signal Processing*, vol. 144, p. 106885, Oct. 2020, doi: 10.1016/j.ymssp.2020.106885.

[64] M. Z. Islam, M. M. Islam, and A. Asraf, “A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images,” *Informatics in Medicine Unlocked*, vol. 20, p. 100412, 2020, doi: 10.1016/j.imu.2020.100412.

[65] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[66] A. Graves, A. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Mar. 2013, pp. 6645–6649, doi: 10.1109/ICASSP.2013.6638947.

[67] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: unified, real-time object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, vol. 2016-Decem, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[68] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A review of Yolo algorithm developments,” *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2021, doi: 10.1016/j.procs.2022.01.135.

[69] J. Redmon and A. Farhadi, “YOLO9000: better, faster, stronger,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, vol. 2017-Janua, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[70] J. Redmon and A. Farhadi, “YOLOv3: an incremental improvement,” *arXiv preprint arXiv:1804.02767*, Apr. 2018, [Online]. Available: http://arxiv.org/abs/1804.02767.

[71] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: optimal speed and accuracy of object detection,” *arXiv preprint arXiv:2004.10934*, Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.10934.

[72] “YOLOv5 documentation,” *Ultralytics*. https://docs.ultralytics.com/#yolov5 (accessed Nov. 03, 2022).

[73] J. Yi, P. Wu, and D. N. Metaxas, “ASSD: attentive single shot multibox detector,” *Computer Vision and Image Understanding*, vol. 189, p. 102827, Dec. 2019, doi: 10.1016/j.cviu.2019.102827.

[74] W. Liu *et al.*, “SSD: single shot multibox detector,” in *Computer Vision – ECCV 2016*, The Netherlands: Springer, 2016, pp. 21–37.

[75] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.

[76] R. I. H. Abushahma, M. A. M. Ali, O. I. Al-Sanjary, and N. M. Tahir, “Region-based convolutional neural network as object detection in images,” in *Proceeding - 2019 IEEE 7th Conference on Systems, Process and Control, ICSPC 2019*, Dec. 2019, pp. 264–268, doi: 10.1109/ICSPC47137.2019.9068011.

[77] R. Girshick, “Fast R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision*, Dec. 2015, vol. 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.
[78] B. Wu, F. Iandola, P. H. Jin, and K. Keutzer, “SqueezeDet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2017, vol. 2017-July, pp. 446–454, doi: 10.1109/CVPRW.2017.60.
[79] T. N. Abu-Jamie and S. S. Abu-Naser, “Classification of sign-language using MobileNet - deep learning,” *International Journal of Academic Information Systems Research (IJAISR)*, vol. 6, no. 7, pp. 29–40, 2022.
[80] I. Goodfellow *et al.*, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.
[81] L. Posilović, D. Medak, M. Subašić, M. Budimir, and S. Lončarić, “Generative adversarial network with object detector discriminator for enhanced defect detection on ultrasonic B-scans,” *Neurocomputing*, vol. 459, pp. 361–369, Oct. 2021, doi: 10.1016/j.neucom.2021.06.094.
[82] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “You only learn one representation: unified network for multiple tasks,” *arXiv preprint arXiv:2105.04206*, May 2021, [Online]. Available: http://arxiv.org/abs/2105.04206.
[83] S. K. Jha, “Advances in object detection algorithms – a brief survey,” *Interantional Journal of Scientific Research in Engineering and Management*, Jan. 2022, doi: 10.55041/ijsrem11467.
[84] J. Kabugo, “Monitoring the waste to energy plant using the latest AI methods and tools,” Kemian tekniikan korkeakoulu, 2018.
[85] R. López, “What is neural designer?,” *neuraldesigner.com*. https://www.neuraldesigner.com/learning/user-guide/what-is-neural-designer (accessed Nov. 02, 2022).
[86] D. Suleiman and G. Al-Naymat, “SMS spam detection using H2O framework,” *Procedia Computer Science*, vol. 113, pp. 154–161, 2017, doi: 10.1016/j.procs.2017.08.335.
[87] P. Schüller *et al.*, “Composing complex and hybrid AI solutions,” *arXiv preprint arXiv:2202.12566*, Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.12566.
[88] A. Tveit, T. Morland, and T. B. Røst, “DeepLearningKit - an GPU optimized deep learning framework for Apple’s iOS, OS X and tvOS developed in Metal and Swift,” *arXiv preprint arXiv:1605.04614*, May 2016, [Online]. Available: http://arxiv.org/abs/1605.04614.
[89] K. Ota, M. S. Dao, V. Mezaris, and F. G. B. De Natale, “Deep learning for mobile multimedia: a survey,” *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 13, no. 3s, pp. 1–22, Jun. 2017, doi: 10.1145/3092831.
[90] M. A. A. Hammoudeh, M. Alsaykhan, R. Alsalameh, and N. Althwaibi, “Computer vision: a review of detecting objects in videos – challenges and techniques,” *International journal of online and biomedical engineering*, vol. 18, no. 1, pp. 15–27, Jan. 2022, doi: 10.3991/ijoe.v18i01.27577.
[91] B. T. Chicho and A. B. Sallow, “A comprehensive survey of deep learning models based on Keras framework,” *Journal of Soft Computing and Data Mining*, vol. 2, no. 2, pp. 49–62, Oct. 2021, doi: 10.30880/jscdm.2021.02.02.005.
[92] S. Aggarwal, M. Bhatia, R. Madaan, and H. M. Pandey, “Optimized sequential model for plant recognition in Keras,” *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 12118, Jan. 2021, doi: 10.1088/1757-899X/1022/1/012118.
[93] B. Pang, E. Nijkamp, and Y. N. Wu, “Deep learning with TensorFlow: a review,” *Journal of Educational and Behavioral Statistics*, vol. 45, no. 2, pp. 227–248, Sep. 2020, doi: 10.3102/1076998619872761.
[94] W. G. Hatcher and W. Yu, “A survey of deep learning: platforms, applications and emerging research trends,” *IEEE Access*, vol. 6, pp. 24411–24432, 2018, doi: 10.1109/ACCESS.2018.2830661.
[95] “ConvNetJS: deep learning in your browser,” *cs.stanford.edu*. https://cs.stanford.edu/people/karpathy/convnetjs/index.html (accessed Nov. 02, 2022).
[96] A. Parvat, J. Chavan, S. Kadam, S. Dev, and V. Pathak, “A survey of deep-learning frameworks,” in *Proceedings of the International Conference on Inventive Systems and Control, ICISC 2017*, Jan. 2017, pp. 1–7, doi: 10.1109/ICISC.2017.8068684.
[97] W. Q. Yan, *Computational Methods for Deep Learning*. Springer Singapore, 2023.
[98] W. Sultani, C. Chen, and M. Shah, “Real-world anomaly detection in surveillance videos,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 6479–6488, doi: 10.1109/CVPR.2018.00678.
[99] R. Jayaswal and M. Dixit, “A framework for anomaly classification using deep transfer learning approach,” *Revue d’Intelligence Artificielle*, vol. 35, no. 3, pp. 255–263, Jun. 2021, doi: 10.18280/ria.350309.
[100] N. R. P. Rodrigues, N. M. C. da Costa, R. Novais, J. Fonseca, P. Cardoso, and J. Borges, “AI based monitoring violent action detection data for in-vehicle scenarios,” *Data in Brief*, vol. 45, p. 108564, Dec. 2022, doi: 10.1016/j.dib.2022.108564.
[101] N. Sikder and A. Al Nahid, “KU-HAR: an open dataset for heterogeneous human activity recognition,” *Pattern Recognition Letters*, vol. 146, pp. 46–54, Jun. 2021, doi: 10.1016/j.patrec.2021.02.024.
[102] J. P. Singh, S. Arora, S. Jain, and U. P. Singh Som, “A multi-gait dataset for human recognition under occlusion scenario,” in *IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques, ICICT 2019*, Sep. 2019, pp. 1–6, doi: 10.1109/ICICT46931.2019.8977673.

BIOGRAPHIES OF AUTHORS

Yunusa Mohammed Jeddah is a Ph.D. scholar at the Department of Electrical and Computer Engineering, International Islamic University Malaysia. He Holds a Master’s degree in computer and information engineering from the same department. His research areas are image/video processing and IoT. His Ph.D. research focuses on recognition of suspicious human behavior using deep learning. His Maters dissertation was on IoT security, using honeypots. He can be contacted at email: yunusmj2@hotmail.com.

Prof. Dr. Aisha Hassan Abdalla Hashim joined IIUM in 1997 and is currently a professor at the Department of Electrical and Computer Engineering. She has been a member of the Department Board of Studies for several years. She received the Best Teacher Award during IIUM Quality Day in 2007. She has been appointed as external examiner/visiting professor/adjunct professor at different universities. She is actively involved in research and postgraduate programmers, has published more than 200 journal/conference papers, and supervised/co-supervised more than 40 Ph.D./Master's students. She has received many medals/awards in different national/international research exhibitions. As a researcher, she has secured research grants from IIUM, Ministry of Higher Education (MOHE) and Ministry of Science, Technology and Innovation (MOSTI). She has been appointed as IIUMÂ Internationalization Ambassador to Sudan (October 2014) and has participated in initiating several MoUs as well as encouraging the Ph.D. Student mobility programmer between IIUM and Sudanese Universities. She can be contacted at email: aisha@iium.edu.my.

Prof. Dr. Othman Omran Khalifa worked in industries for eight years and he is currently a professor at the Department of Electrical and Computer Engineering, International Islamic University Malaysia. He served as the head of Department of Electrical and Computer Engineering, IIUM from July 2005 until December 2014. He is a Charter Engineer (CEng) UK, a senior member of IEEE USA, and a member IET UK. He was the chairman of the International Conference on Computer and Communication Engineering (ICCCE), 2006, 2010, 2012, and 2014. He supervised more than 60 masters and Ph.D. students. He has extensively contributed through his writings in international journals, conferences and books. He published more than 500 publications including 12 books. In 2013, he received the highest citation in Citation Indexed Journal Award, in 2019 he got the TOKOH academic “Best Academician” award and last year in 2020, he won the Murabbi “Best Teacher” award. In 2021, he got the publication (book publication) award. In 2022 won the National Best Book in Science and Technology 2022. He can be contacted at email: khalifa@iium.edu.my.