Computers in Biology and Medicine 137 (2021) 104803

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Artificial intelligence-based hybrid deep learning models for image classification: The first narrative review

Biswajit Jena [a], Sanjay Saxena [a], Gopal K. Nayak [a], Luca Saba [b], Neeraj Sharma [c], Jasjit S. Suri [d,*]

[a] Department of CSE, International Institute of Information Technology, Bhubaneswar, India
[b] Department of Radiology, University of Cagliari, Italy
[c] School of Biomedical Engineering, IIT (BHU), Varanasi, India
[d] Stroke Diagnosis and Monitoring Division, AtheroPoint™, Roseville, CA, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Deep learning
Hybrid deep learning
Spatial
Temporal
Spatial-temporal
Performance
Risk-of-bias

ABSTRACT

Background: Artificial intelligence (AI) has served humanity in many applications since its inception. Currently, it dominates the imaging field—in particular, image classification. The task of image classification became much easier with machine learning (ML) and subsequently got automated and more accurate by using deep learning (DL). By default, DL consists of a single architecture and is termed solo deep learning (SDL). When two or more DL architectures are fused, the result is termed a hybrid deep learning (HDL) model. The use of HDL models is becoming popular in several applications, but no review of these uses has been designed thus far. Therefore, this study provides the first narrative HDL review by considering all facets of image classification using AI.

Approach: Our review employs a PRISMA search strategy using Google Scholar, PubMed, IEEE, and Elsevier Science Direct, through which 127 relevant HDL studies were considered. Based on the computer vision evolution, HDLs were subsequently classified into three categories (spatial, temporal, and spatial-temporal). Each study was then analyzed based on several attributes, including continent, publisher, hybridization of two DL or ML, architecture layout, application type, data set type, dataset size, feature extraction methodology, connecting classifier, performance evaluation metrics, and risk-of-bias.

Conclusion: The HDL models have shown stable and superior performance by taking the best aspects of two or more solo DL or fusion of DL with ML models. Our findings indicate that HDL is being applied aggressively to several medical and non-medical applications. Furthermore, risk-of-bias is highly debatable for DL and HDL models.

1. Introduction

Artificial intelligence (AI) represents an innovative breakthrough that has paved the way for various applications in different domains with great success since its inception [1]. It also forms the basis for all computer learning and has supported complex decision-making [2–4]. Moreover, the cutting-edge technologies associated with AI have made our lives more comfortable and serve humanity better as they become more precise [5,6]. AI has become an integral part of our daily lives, whether we are reading emails [5], getting driving directions [6], playing music, getting movie recommendations [7], using self-driving and self-parking vehicles [8], or talking with chat-bots [9]. AI has now even become central to environmental changes while supporting pandemic-related challenges [10–13].

AI has become smarter with the innovation of deep learning (DL) [14,15]. DL is a subset of ML by which artificial neural networks (Appendix B: Figure B1) and algorithms inspired by the human brain learn from input data [15–17]. The role of AI has deeply influenced the imaging science domain, particularly regarding image classification in big data frameworks. The process has become automatic with the introduction of AI in classification tasks, thereby allowing AI-based systems to achieve human-level performance [14]. AI is an integral part of image processing that classifies a wide range of images into their respective classes. As DL learns automatically, like the human brain, image classification tasks are becoming smoother via an automatic feature extraction paradigm within a big data framework [14,15,18,19].

DL models based on single DL architecture are termed solo deep learning (SDL) models. When two DL architectures are connected by a bridge network, or when DL is cascaded with ML for classification

* Corresponding author. Stroke Diagnosis and Monitoring Division, AtheroPoint™ LLC, Roseville, CA, USA.
E-mail address: jasjit.suri@atheropoint.com (J.S. Suri).

https://doi.org/10.1016/j.compbiomed.2021.104803
Received 27 June 2021; Received in revised form 22 August 2021; Accepted 23 August 2021
Available online 27 August 2021

Nomenclature

AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area-under-the-curve
CNN	Convolutional Neural Network
CT	Computer Tomography
BiLSTM	Bidirectional Long Short-Term Memory
DBN	Deep Belief Network
DL	Deep Learning
EGG	Electroencephalogram
GRU	Gated Recurrent Unit
HAR	Human Activity Recognition
HDL	Hybrid Deep Learning
IDS	Intrusion Detection System
HSI	Hyperspectral Image
KNN	K-Nearest Neighbour
LRL	Linear Regression Layer
LSTM	Long Short-Term Memory
ML	Machine learning
MRI	Magnetic Resonance Imaging
MI	Motor Imagery (Brain)
NLP	Natural Language Processing
NN	Neural Network
OCT	Optical Coherence Tomography
PE	Performance Evaluation
PET	Positron Emission Tomography
PSO	Particle Swarm Optimization
RF	Random Forest
RNN	Recurrent Neural Network
RBM	Restricted Boltzmann Machine
SRU	Simple Recurrent Unit
SDL	Solo Deep Learning
SVM	Support Vector Machine
TL	Transfer Learning
VGG	Visual Geometric Group
WOA	Whale Optimization Algorithm

purposes, the model is categorized as a hybrid deep learning (HDL) model [16,20,21]. This hybridization process is often referred to as the fusion process or the process of combining two or more SDL architectures. More specifically, we can say that HDL deals with various kinds of inputs (images, videos, and electronic time-series signals) with multiple AI-based architectures referenced in this study [10–13,22–78].

Typically, HDLs are application-oriented, but they frequently try to obtain better performance. The basic differences between SDL and HDL are shown in Table 1. The most popular HDL applications are in the area of healthcare [10,11,13,23,28,37,38,45,52,55,62,64,70,74], hyperspectral image classification [27,30,47,48,53,69,71,73], audio-visual emotion recognition [36,46,68,75,76], human activity recognition [12,25,26,31,34,35,44,57,60], time-series data analysis [41,58], traffic flow analysis, and video surveillance [33,39,40,66,67].

From some of our initial analyses of the HDL studies [10,11,13,62,64], it is apparent that HDL performs better than SDL for the same dataset. Due to the inherent advantages of HDL over SDL, we hypothesize that HDL is superior to SDL models for classifying or characterizing the processes leading to classification. Appendix A shows the difference between HDL and SDL at the performance level. We designed our fundamental HDL classes in line with the computer vision field's evolution from spatial to temporal and from hopping to spatial-temporal, enveloping all applications.

Szegedy et al. [79] were the first to coin the term HDL, which basically consisted of combining two fundamental SDL models such as Inception and ResNet, leading to an HDL model known as Inception-ResNet. Before this, the hybridization process was limited to the fusion of DL and ML paradigms [42,50,56]. Since 2016, the hybridization process has accelerated, both for homogeneous (similar architecture) and non-homogeneous (heterogeneous) architecture designs. Fig. 1 shows the trend of HDL studies over the last ten years.

Even though the number of articles focused on HDL has increased linearly, no narrative review of such articles has been published, except one recent specialized publication in 2021 on a brain-computer interface application [16].

This narrative review, which selects the best **127** publications using the PRISMA model, is the first of its kind. This review presents (i) the fundamental classes of HDL models with clear examples and results, (ii) architectures of these fundamental HDL classes, and (iii) a comprehensive statistical analysis that considers various AI attributes, such as feature extraction strategy, classifier used, performance evaluation metrics, optimizer, loss function, and learning rate.

2. Search strategy and statistical distributions of HDL models

2.1. PRISMA model design

A detailed search was performed using Google Scholar, PubMed, IEEE Xplore, ScienceDirect, and arXiv. The keywords used for selecting studies were "hybrid deep learning", "hybrid deep learning review", "hybrid models", "spatial classification using the hybrid model", "temporal classification using the hybrid model", and "spatial-temporal classification using the hybrid model", and "fusion of deep learning models". Fig. 2 shows the PRISMA model consisting of the HDL references used in this study.

After an exhaustive search, a total of 190 studies were identified; duplicates were removed using the "Find Duplicates" feature in EndNote software by Clarivate Analytics [80]. After this process, 177 records were retained. The three exclusion criteria were (**i**) studies not related to AI, (**ii**) non-relevant articles, and (**iii**) articles with insufficient data. After applying the exclusion criteria, 23, 14, and 13 studies (marked as E1, E2 (non-AI, but hybrid), and E3 in Fig. 2) were identified, leading to

Table 1
Solo deep learning *vs.* hybrid deep learning.

Attributes	SDL	HDL	References
Feature Extraction	Limited scope	Larger scope	[10–13,22–79]
Classifier Diversity	Limited to Softmax	Softmax or ML-based	[10,43,62,65]
Transfer Learning	Limited options for transfer learning	More options for transfer learning	[10,11,13,28,38,43,55,57,60,62,64]
PE Value	Low performance	Superior performance	[10,11,13,62,64]
Hardware Resource	Uses low resource	Uses more resources	[36,51,64,65]
Program Complexity	Low complexity	High complexity	[10,11,13,62,64]
Applications	Limited application	Diverse application	[12,36,46–48,53,60,66,67,74–76]

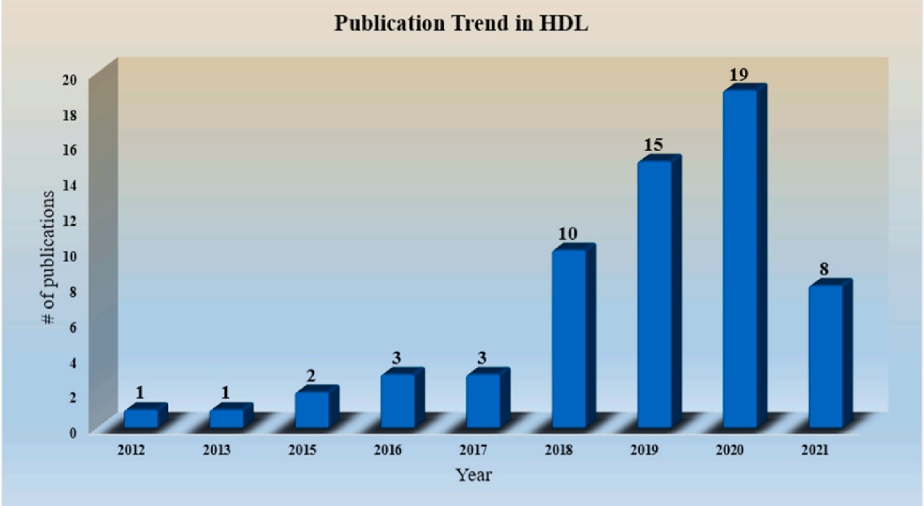

Fig. 1. Number of HDL publications per year over the last ten years.

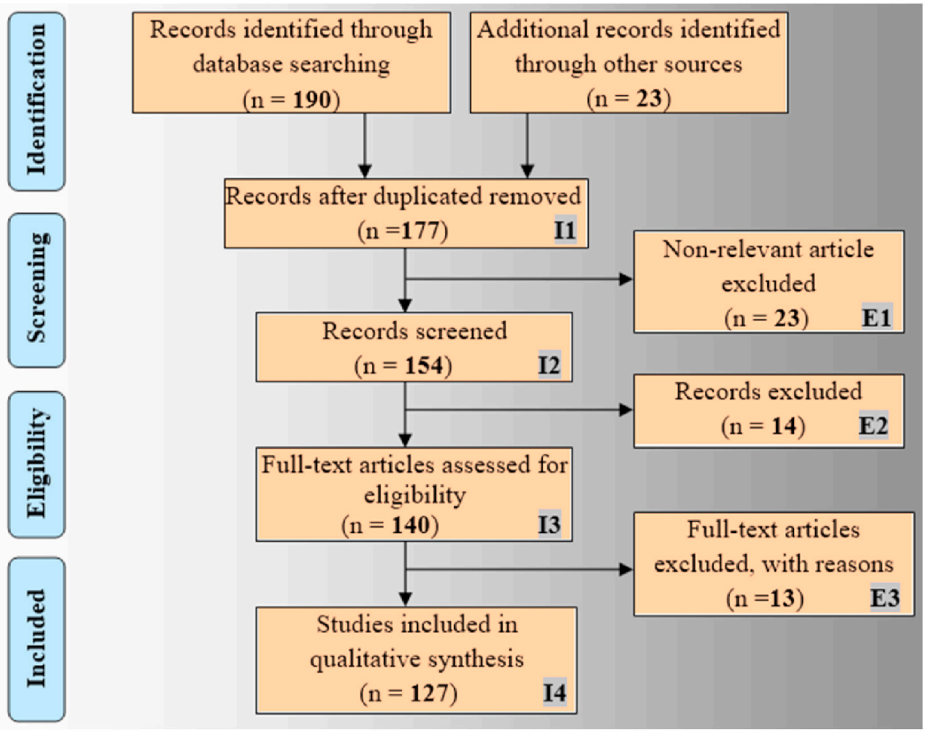

Fig. 2. PRISMA analysis for selection of studies.

the final selection of **127** references for this study.

It is essential to extract information from these studies and understand their distributions statistically. These matters are discussed in the following section.

2.2. Statistical distributions

2.2.1. Distribution of HDL publications by Publisher and Geographical Continents

As discussed in the introduction, the HDL models examined in this study have been classified into three fundamental types: (i) spatial, (ii) temporal, and (iii) spatial-temporal. When the data comes from static images, the model belongs to the spatial HDL (sHDL) class. Similarly, when the data are taken from the temporal domain, the model belongs to the temporal HDL (tHDL) class. Finally, if the data consist of a combination of static and temporal information, the model belongs to the spatial-temporal HDL (stHDL) class. The models' categorization depends on the nature of the input modalities entered into the HDL model and the structure (or stages) of the HDL model. The details of the architecture are explained in Section 3.

The spatial class [10,11,13,22–24,28,29,33,37,38,40,42,43,45,50–52,55,56,59,62,64,65,72] contributed **41%** of the HDL cohort, the temporal class [25,31,32,34,35,41,49,54,58,63,69,77,78] covered **22%**, and the spatial-temporal class [12,26,27,30,36,39,44,46–48,53,57,60,66–68,70,71,73–76] covered **37%** of the cohort (Fig. 3 (a)).

With evolution of HDL models throughout the world, it is important to understand the origin of HDL by publishers and continents. For that purpose, our HDL studies were further distributed according to publishers and

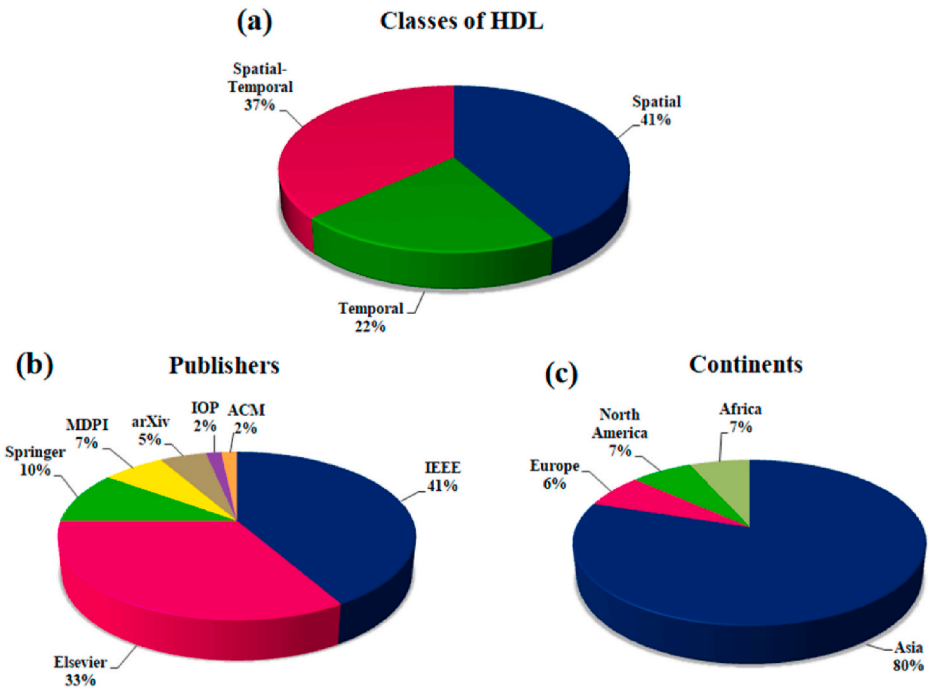

Fig. 3. HDL studies distribution by (a) HDL type, (b) publisher, and (c) continent.

continent of origin. Fig. 3 (b) presents a pie chart showing that the most common publisher is IEEE (**41%**) [10,12,13,29–31,33,36,37,39,41,44,48,53,54,56,58,60,63,65,71,73,75–77], followed by Elsevier (**33%**) [11,23,26,32,34,35,43,45,46,49–51,55,57,62,67,69,70,72,74], Springer (**10%**) [25,28,38,42,47,64], Multidisciplinary Digital Publishing Institute (MDPI) (**7%**) [27,40,52,78], arXiv (**5%** [22,59,66], Institute of Physics (IOP) (**2%**) [24], and Association for Computing Machinery (ACM) (**2%**) [68].

Fig. 3 (c) shows the HDL publications by continent. The greatest contributors were from Asia (**60%** of the world's population) (**80%**) [12,13,23–27,30–34,36–42,44–48,51–57,60,62–68,70–78]. North America [29,50,58,59], Africa [10,11,35,43], and Europe [22,28,49,69] each contributed ~**7%**.

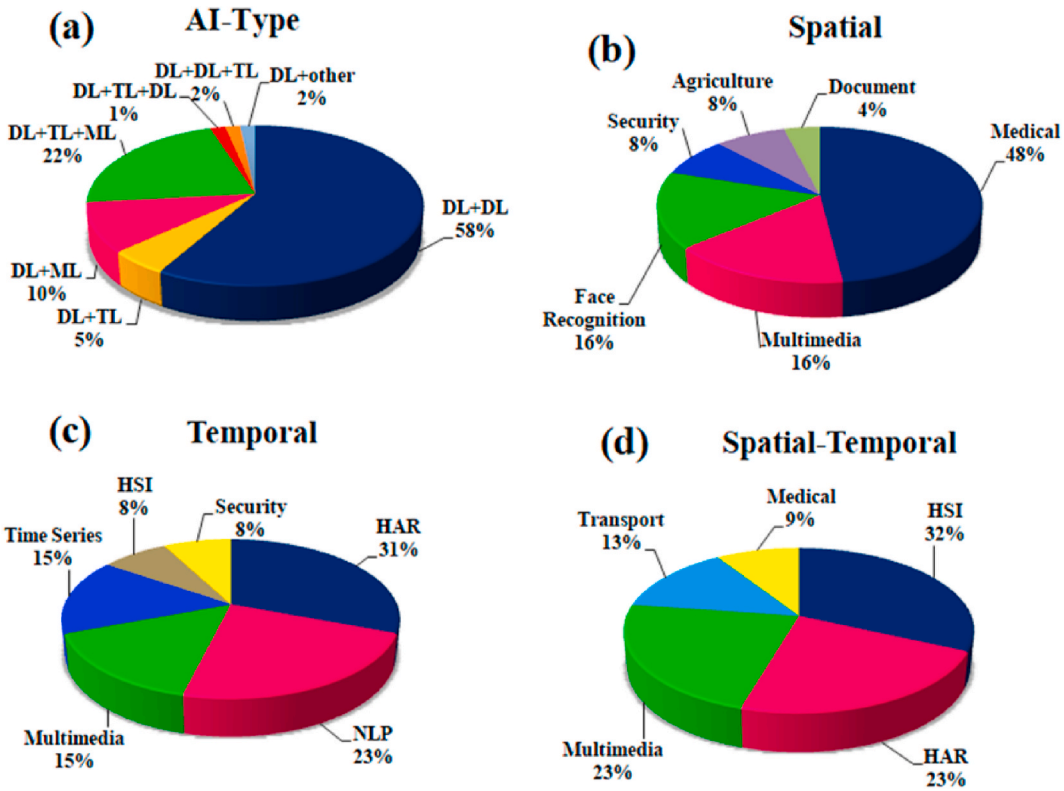

Fig. 4. Statistical distribution by (a) AI type and (b) HDL application by (i) spatial (ii) temporal, and (iii) spatial-temporal type.

2.2.2. Types of hybridization using different kinds of ML/DL/TL combinations

HDL models are combinations of various AI-based models intended to develop a robust DL model. The various AI-based models used for this purpose were deep learning (DL), machine learning (ML), transfer learning (TL), and combinations of AI optimization technologies [10,69]. Based on these, Fig. 4 (a) shows various combinations of AI types.

The major AI types were "DL + DL" (**58%**) [12,22,23,25–27,30–37,39–41,44,47,49,51,53,54,56,58,59,63,66–68,70,71,73,77,78], followed by "DL + TL + ML" (**22%**) [10,11,13,24,43,46,55,60,62,64,65,75,76], "DL + ML" (**10%**) [29,42,48,50,52,72], "DL + TL" (**5%**) [28,38,45], "DL + DL + TL" [74], "DL + others" (**2%**) [69], and "DL + TL + DL" (**1%**) [57]. Note that these AI type categories were grouped into three types of HDL classes: spatial, temporal, and spatial-temporal.

2.2.3. Distribution of application types in each of the three fundamental HDL classes

A wider variety of applications has been found under the umbrella of HDL. In some scenarios, the HDL has been designed to handle specific applications. The broad applications of DL are solvable by HDL models, which showed superior performance to other models. The applications of HDL models are again classified based on the three classes of such models. However, some applications belong to multi-classes of HDL.

Fig. 4 (b), 4 (c), and 4 (d) show various HDL applications for the three distinct HDL classes. The major applications covered under HDL are medical, multimedia, human action recognition (HAR), natural language processing (NLP), hyperspectral image (HSI) analysis, time-series data analysis, and traffic flow analysis. "Spatial class" models are shown in Fig. 4 (b). The largest group of these are medical (**48%**) [10,11,13,23,28,37,38,45,52,55,62,64], multimedia (**16%**) [22,33,40,59], face recognition (**16%**) [24,43,51,56], security (**8%**) [29,72], agriculture (**8%**) [42,65] and document (**4%**) [50].

The applications under "temporal class" are shown in Fig. 4 (c). The contributors to this class are HAR (**31%**) [25,31,34,35], NLP (**23%**) [54,77,78], multimedia (**15%**) [49,63], time-series (**15%**) [41,58], HSI (**8%**) [69], and security (**8%**) [32]. Finally, the "spatial-temporal" class (Fig. 4 (d)) consisted of hyperspectral image (HIS) (**32%**) [27,30,47,53,71,73], HAR (**23%**) [12,26,44,57,60], multimedia (**23%**) [36,46,68,75,76], transport (**13%**) [39,66,67], and medical (**9%**) [70,74].

2.2.4. Distribution of data size used in three fundamental HDL classes

Data size was one of the key factors in deciding how robust the HDL model is. If the model is trained with sufficient data and tested using an adequate number of data sets, that model is robust and can be free from over-fitting and data imbalance problems. The top 10 studies in each of the three HDL classes are shown in Fig. 5.

By dataset size, we mean the number of image and video files taken of the modalities (natural images, CT, PET, MRI, X-ray, HSI, surveillance videos, and time-series images). The top 10 data sizes for the spatial, temporal, and spatial-temporal classes were contributed by Refs. [22–24,38,43,50,51,56,64,72], [25,31,32,34,49,54,58,63,77,78], and [12,36,57,66–68,70,75], respectively.

2.2.5. Distributions of input imaging modalities in three fundamental HDL classes

The principle of classifying the whole set of HDL models into three HDL classes was governed by the status of each model's input modality. As defined earlier, the spatial class consisted of input modality: static, time-series data for the temporal class, and a combination of the static-and-temporal data for the spatial-temporal HDL class. Another category of image data consisted of document and text images. The complete input modalities of the HDL model are depicted in Fig. 6 (a).

Natural images represent the largest type of data input to the HDL model (**28%**) [22,24,25,28,33–35,40,42,43,50,51,56,57,59,65,67], followed by HSI (**13%**) [27,30,47,48,53,69,71,73], as both temporal and spatial-temporal data are controlled by HSI. The video and HSI are considered in temporal and spatial-temporal classes of HDL models, contributing **22%** [26,31,36,44,46,60,66,68,75,76], and **13%** [27,30,47,48,53,69,71,73], respectively. The medical applications handled by HDL models are controlled by various medical imaging modalities, such as CT (**7%**) [10,11,52,62], PET (**2%**) [55], MRI (**3%**) [38,45], X-ray (**3%**) [13,23], OCT (**2%**) [37], MI (**2%**) [74].

Similarly, the input modalities of spatial, temporal, and spatial-temporal classes (and their proportions) are shown in Fig. 6 (b), 6 (c), and 6 (d). For the spatial class, the top two contributors were natural images and CT. For the temporal HDL class, the top two contributors

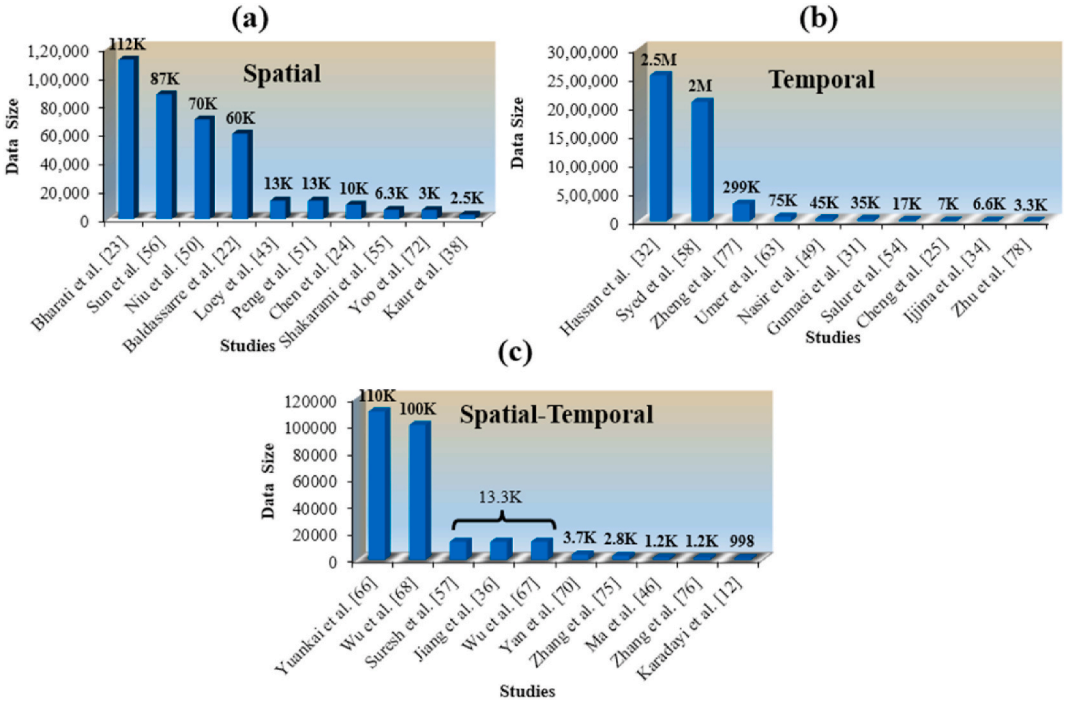

Fig. 5. Top 10 studies for each of the three HDL classes: spatial, temporal, and spatial-temporal.

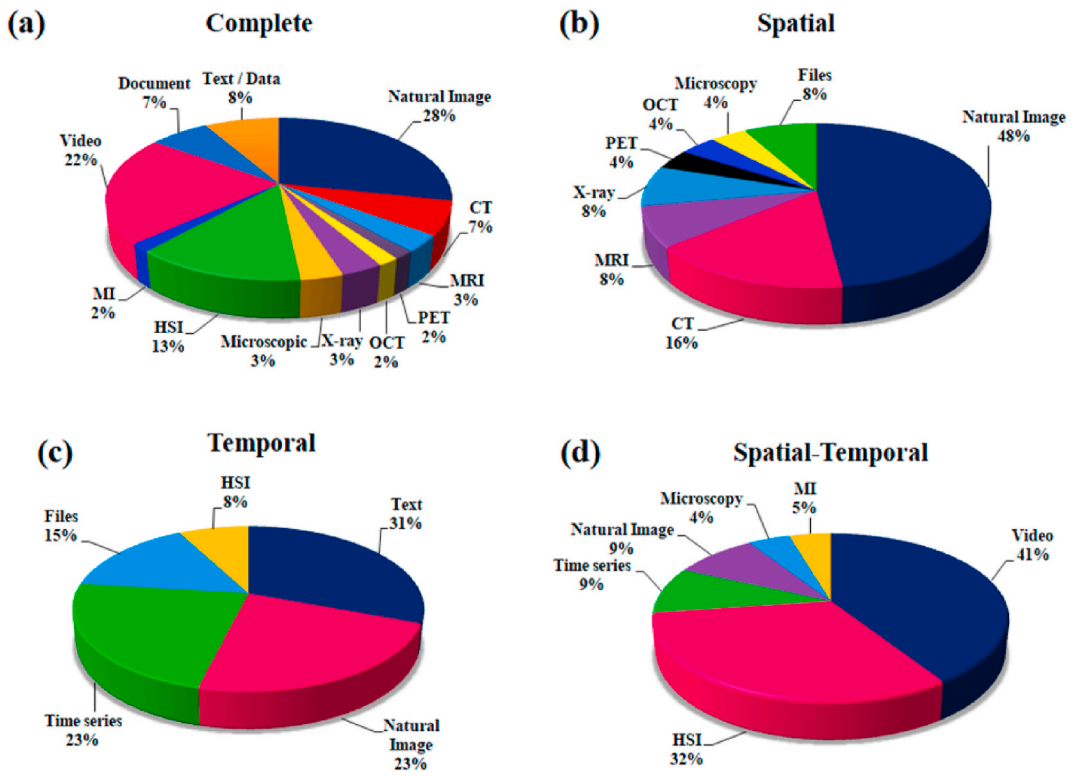

Fig. 6. Various types of input modalities adapted in three kinds of HDL classes.

were natural images and time-series. For the spatial-temporal HDL class, the top two contributors were videos and HSI.

3. Introduction to HDL architectures

Scene processing in computer vision began with static imagery and evolved into motion imagery [81]. The detection of objects among scenery has been fundamental to image processing in static imagery applications [82]. The invention of the video camera and the subsequent availability of videos led to the processing of motion images [61]. With applications of high-speed hardware such as CPUs and GPUs, alongside the evolution of superior feature detection methods using AI, spatial and temporal information fusion started to evolve [46,75,76]. In the spirit of computer vision, AI-based deep learning models also started to evolve in three HDL classes using imaging and non-imaging science spectra. Thus, HDL-based learning models are categorically arranged into these three classic imagery paradigms. HDL models are now used to solve challenges related to the classification of scenes under different conditions (e.g., environmental changes, pandemic occurrences [10,11,13], sports, HAR [12,25,26,31,34,35,44,57,60], and bioinformatics). They are now a routine part of the static or spatial domain, the tracking overtime or temporal domain, and the fusion of spatial and temporal information (the so-called spatial-temporal domain) [36,46,68,75,76].

3.1. Spatial HDL architecture

The spatial HDL architecture is characterized by input modalities based on spatial or static imagery. The hybridization process also plays an important role in deciding the spatial HDL architecture. All 2D CNNs are suitable deep learning models for analyzing the spatial characteristics of an image. This includes state-of-the-art CNN models like AlexNet [83], Visual Geometry Group (VGG) from Oxford [84], ResNet [85], and GoogleNet [79]. These models were fused during the hybridization process to design the spatial HDL architecture that supports classification. The first stage of spatial HDL is the DL architecture, which is mainly used for feature extraction. The second stage can be either a DL or ML system, typically adapted for classification tasks [10,11,13,22–24,28,29,33,37,38,40,42,43,45,50–52,55,56,59,62,64,65,72]. In some scenarios, transfer learning is part of the hybridization process to form spatial HDL architecture—in such scenarios, pre-trained weights are used as feature extraction (stage one), followed by a DL- or ML-based classifier (stage two) [11,13,55,62,64].

Some high-performing spatial architectures are typically chosen over others in our study. Rezaee et al. [13] attempted to develop a multi-class framework to classify four kinds of pneumonia using an HDL model that hybridizes various popular pre-trained CNN models (such as AlexNet, VGG, and CaffeNet) (as part of stage one) and an ML-based support vector machine (SVM) classifier (as part of stage two) (Fig. 7). Their multi-4 class classification HDL model yielded an accuracy of 99.5%. Extracting high-quality features from the pre-trained CNN models and subsequently passing them on to ML classifiers such as SVM provides sufficient power to design a highly accurate HDL model. When combined with ML, this HDL model provides the flexibility required to improve the overall system's performance. Perhaps the most beneficial aspect of this spatial HDL architecture is combining three similar pre-trained SDL models with nearly similar architectures to extract high-quality features, followed by a soft-voting method [13] for feature selection during hybridization.

Elkorany et al. [11] also provided a multi-class (with four class paradigms) COVID application using ShuffleNet and SqueezeNet (for stage one), combined with an ML-based SVM classifier (stage two) (Fig. 8). They obtained an accuracy of 94.45%. The most notable benefit of this HDL architecture is the use of improved global average pooling layers as part of its feature extraction methodology. The essential details about these two architectures are given in Table 2. These two spatial HDL class architectures are simple and utilize a combination of transfer learning and an SVM classifier to yield high performance. Furthermore, these architectures include high-performing implementation settings to deal with the risk-of-bias analysis (presented later) and avoid over-fitting.

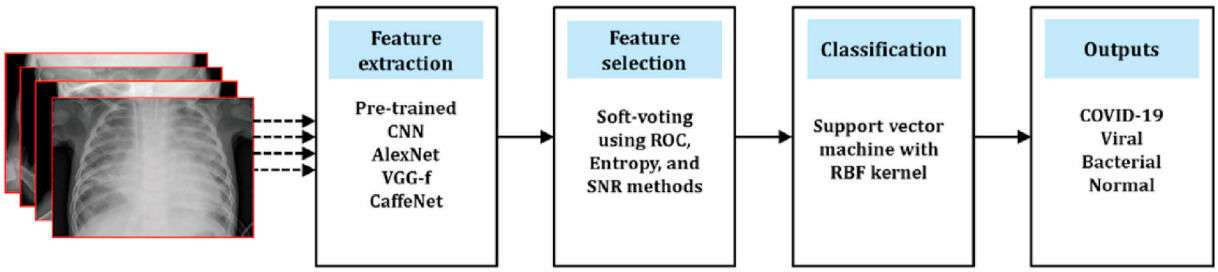

Fig. 7. Hybridization of various CNN models like AlexNet, VGG, and Caffenet with an ML-based SVM classifier [13] to classify four kinds of pneumonia using the multi-class framework (reproduced with permission).

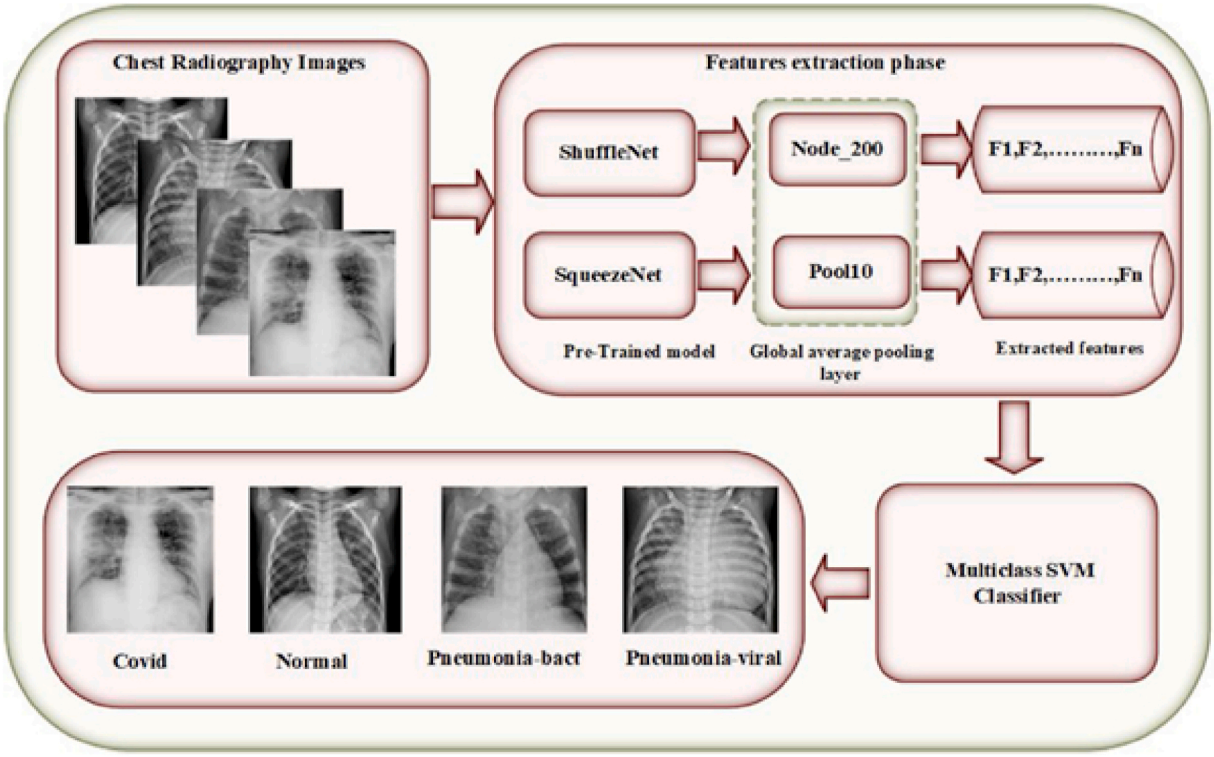

Fig. 8. Hybridization of various CNN models such as ShuffleNet and SqueezeNet (stage one) combined with ML-based SVM classifier (stage two) for COVID application [11] (reproduced with permission).

Table 2
Three types of HDL architectures.

Classes of HDL	SN	HDL Architecture	Stage 1	Stage 2	#CL	#FCL	Result (ACC[a])
Spatial	1	AlexNet, CaffeNet, VGG + SVM [13]	AlexNet, CaffeNet, VGG	SVM	5	3	99.5
	2	COVIDetection Net [11]	ShuffleNet, SqueezeNet	SVM	2	1	94.45
Temporal	3	ML + GRU [35]	CNN	GRU (Softmax)	_	1	96.3
	4	SRU + GRU [31]	SRU	GRU (Softmax)	_	1	99.8
Spatial- Temporal	5	CNN + DBN + SVM [75]	CNN + DBN	SVM	_	_	71.43
	6	2D-CNN+3D-CNN + LSTM [36]	2D-CNN+3D-CNN + LSTM	Softmax	_	_	90.3

aACC: accuracy, SRU: simple recurrent unit, +: Hybrid; #CL: number of convolution layers; #FCL: number of fully connected layers.

3.2. Temporal HDL architecture

The temporal architecture is characterized by temporal input modalities such as 3D images, video, and time-series data. The HDL hybridization process also plays an essential role in deciding the temporal HDL architecture. Since the temporal HDL architecture inputs temporal image modalities, the best DL architectures used in this HDL model are recurrent neural networks (RNNs), such as long short-term memory (LSTM) (Appendix B: Figure B2) [32,49,58,63], bi-directional LSTM [54,69,77,78], gated recurrent unit (GRU) (Appendix B: Figure B3) [31, 35], and simple recurrent unit (SRU) [31]. Also, 3D CNN [34] and deep belief networks such as the restricted Boltzmann machine (RBM) [41]

have been used in some architectures.

In the hybridization process, stages one and two of the DL architectures are used for feature extraction and classification tasks. This type of HDL model deals with temporal feature extraction and, subsequently, classification paradigms. The temporal feature may be extracted by 3D CNN, RNN, or TL (using pre-trained weights). However, the classification is mainly handled by the RNN or LSTM, BiLSTM, GRU, and SRU.

Fig. 9 provides a temporal HDL architecture for HAR that uses ML-based unsupervised methods for motion tracking (stage one) with GRU for action recognition (stage two). Here, ML-based unsupervised methods—such as the Gaussian mixture model (GMM) and Kalman filter (KF), used for motion tracking—provide intense and complex features. The presence of GRU with increased computation powers is then adopted for sequential data and video classification. This HDL design approach is suitable for HAR application, as it uses the best DL architecture (GRU) for temporal images and achieves a high accuracy of 96.3%.

Fig. 10 shows another architecture for HAR using hybrid SRU-GRU with a high accuracy rate of 99.8%. SRUs' simplicity and speed make them advantageous, while GRUs are powerful. Fig. 10 shows the complete architecture of HAR, which has several components.

Various wearable body sensors are placed on the patients to capture the framework's input data and to record the multimodal raw data of their activities' signals. The first component of the figure contains the reshaping phase, during which the signals are processed as channels. Every channel represents a class of activity. The second component of the figure (stage one of HDL) is the deep SRUs-GRUs neural network model, which consists of four hidden layers in addition to the input and output layers. These components are followed by a fully connected layer with a Softmax activation function (stage two of HDL), which represents the final classification and initiates the activity.

This HDL design approach is also suitable for HAR application as it uses the best hybrid DL architecture (SRU-GRU) for temporal images and achieves a high accuracy of 99.8%. Furthermore, these two discussed HDL temporal architectures include high-performing implementation settings, thus avoiding the risk-of-bias analysis (presented ahead) and preventing over-fitting.

3.3. Spatial-temporal HDL architecture

The spatial-temporal architecture is characterized by the model's spatial and temporal input modalities, such as audio-visual, video, HSI, traffic-flow images, and time-series data. This architecture class can solve broad categories of problems since it offers unique advantages for each input modality and its architecture. The HDL stages also play an important role in deciding the complexity of the architecture. As the spatial-temporal architecture takes input both from the spatial and temporal modalities, the most suitable DL models used in the spatial-temporal HDL model during stage one are all conventional 2D-CNN models [27,30,36,47,53,68,71,73,75,76]. In stage two, the most suitable models are RNNs, such as LSTM [12,26,36,39,44,57,60,66–68,70,74], bi-directional LSTM, GRU, 3D-CNN [12,27,30,36,47,53,68,71,73,75,76], and deep belief networks such as (RBM) (Appendix B: Figure B4) [46,48,75,76].

In another scenario, both spatial-temporal features may be extracted during stage one by combining all traditional 2D-CNN models and 3D-CNN or RNN or TL pre-trained weights. Again, the classification process is handled during stage two by 2D-CNN, RNNs (LSTM, BiLSTM, GRU, and SRU), and 3D-CNN.

Fig. 11 shows the architecture for facial expression recognition in video sequences, yielding a classification accuracy of 71.43%. This architecture employs two individual CNNs—including a spatial CNN processing static facial images and a temporal CNN processing optical flow image—to learn high-level spatial and temporal features separately on the divided video segments. Then, the DBN is used as a fusion network to collect spatial and temporal features and classify them using SVM during stage two. This kind of spatial-temporal HDL architecture design can be considered as the most stable architecture of this type, as it handles spatial and temporal features separately, followed by a fusion network with a classifier.

Fig. 12 depicts the hybrid model with three CNNs operating on video frames, stacked optical flow images, and audio signals for spatial, motion, and audio features, respectively. The model achieves an accuracy of 90.3% for video classification.

This architecture is a standard and stable spatial-temporal HDL architecture design. Stage one of this HDL architecture includes both 2D-CNN (spatial feature) and 3D-CNN (temporal feature) and a fusion network (LSTM) for combining the spatial and temporal features. The classification task is subsequently performed during stage two of this architecture.

There are no specific rules for determining the number of layers to use in these kinds of HDL architectures. Nevertheless, the main objective is to minimize the number of layers to make the architecture compact and reduce the number of network parameters to reduce its computational complexity. As a result, these two HDL paradigms of the spatial-temporal architecture can be categorized as high-performing designs—since the likelihood of over-fitting is low, generalization is possible with a low risk-of-bias.

3.4. Statistical distribution of HDL attributes

The various essential components of HDL models that make them robust are summarized below and their statistical distributions.

3.4.1. Types of feature extraction in the HDL framework

Feature extraction is vital to any classification model. The algorithms or models that include the best feature extraction methods perform classification tasks very well. With the invention of DLs and HDLs, feature extraction methods have become automatic and better than ML-based handcrafted methods. HDL models are fusions of DL, ML, TL, and other methodologies. Thus, HDL models have more options for feature extraction than any of the SDLs included in them. HDL models can use automated CNN-based [86], handcrafted ML-based [87–90], TL-based using pre-trained weight [91], and other mixed-methodology feature

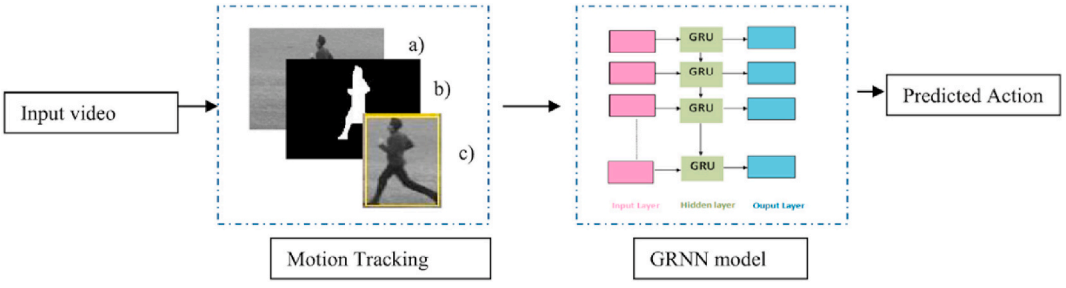

Fig. 9. Hybridization of an unsupervised ML-based method for motion tracking (stage one) with GRU for human action classification (stage two) [35] (reproduced with permission).

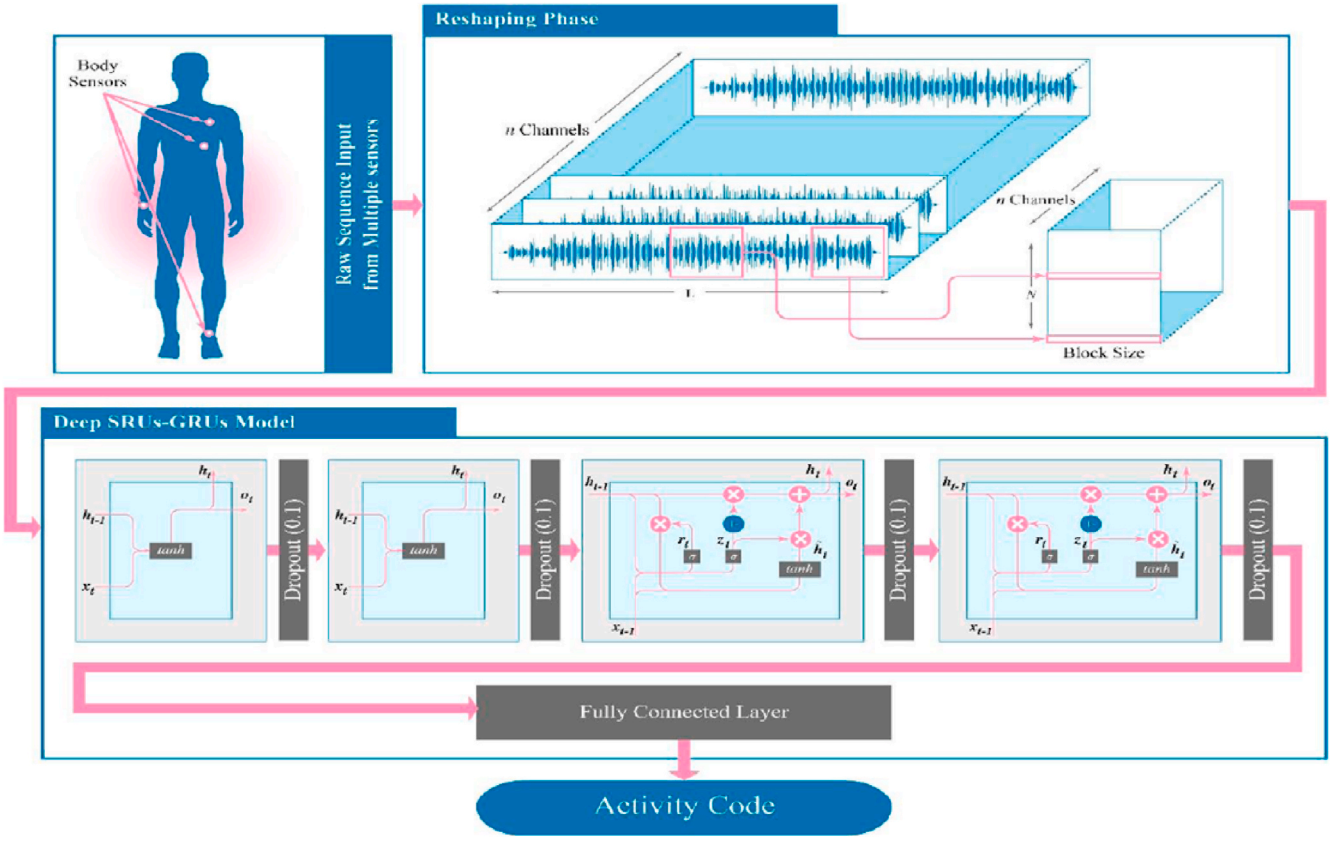

Fig. 10. Example 2 of temporal HDL: Hybridization of SRU and GRU for HAR [31] (reproduced with permission).

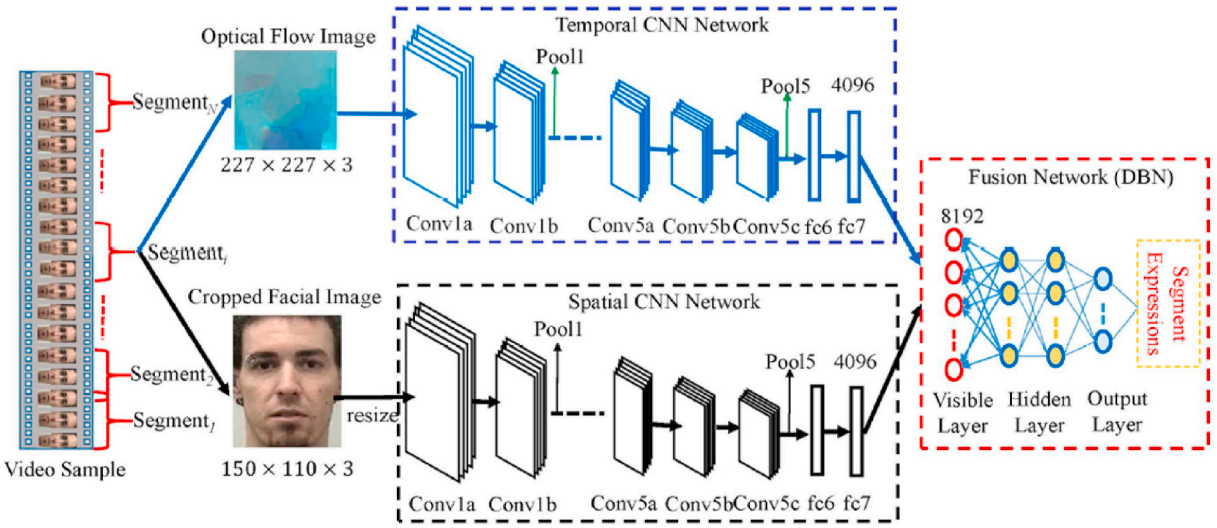

Fig. 11. Hybridization among 3D CNN [stage one] (top in blue dotted rectangle), 2D CNN (bottom part in black dotted rectangle), and DBN [stage two] for facial expression recognition in video sequences [75] (reproduced with permission).

extraction paradigms. In this study, both CNN-based and TL-based [10,11,55,62,64] methods are considered automatic feature extraction methods. The feature extraction strategy in all three classes of HDL concluded that **93%** of the studies adopted the automated feature extraction method, while only **7%** [34,35,72] utilized non-automated methods.

Fig. 13 (a) depicts the distribution of the feature extraction strategies applied by all the HDL model studies. Specifically, **93%** of models are automated, **2%** [72] are handcrafted, and **5%** [34,35] are categorized as "other." A different approach for gathering spatial features from COVID-19 CT images (i.e., a CNN-based global average pooling approach [11]) is used.

This approach calculates the average output of each feature map in the previous layer. This reasonably simple operation reduces the data significantly and prepares the model for the final classification layer, thereby reducing the number of parameters in the model and preventing over-fitting. In a previous study, researchers used a special approach to capture intense and complex temporal features via unsupervised

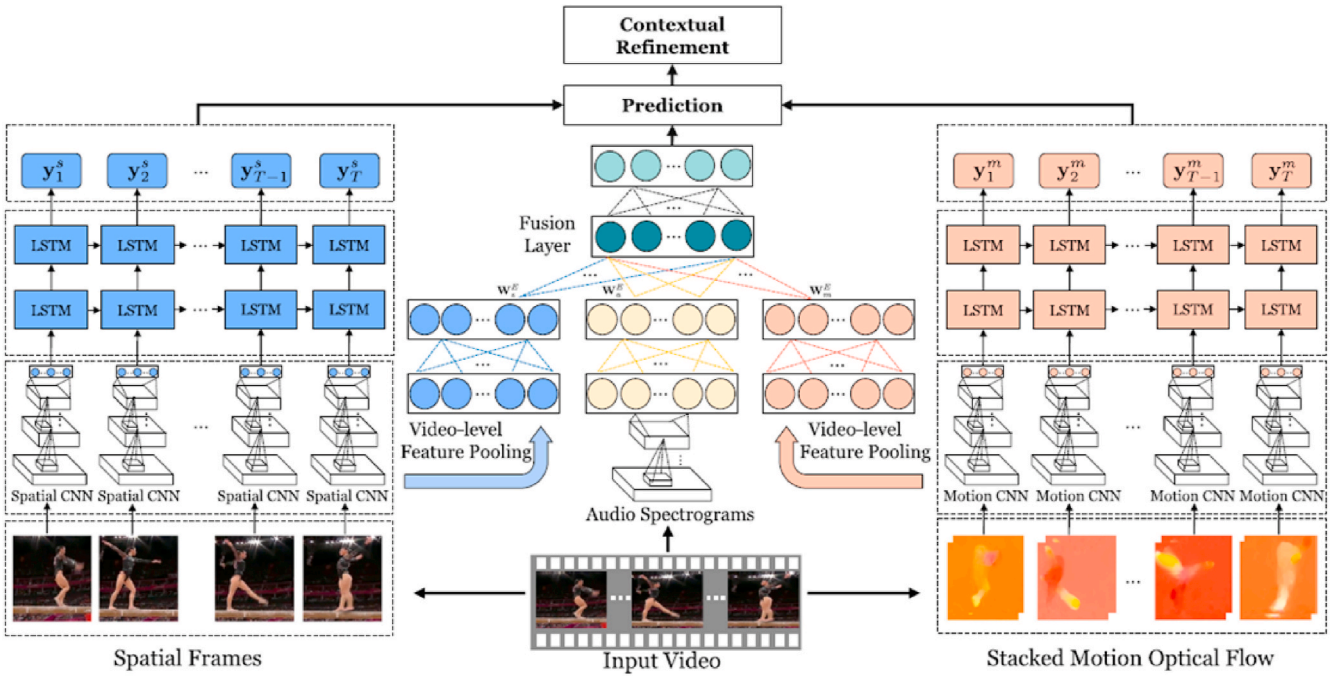

Fig. 12. Hybridization among 2D-CNN, 3D-CNN (stage one), and LSTM (stage two) for video classification [36] (reproduced with permission).

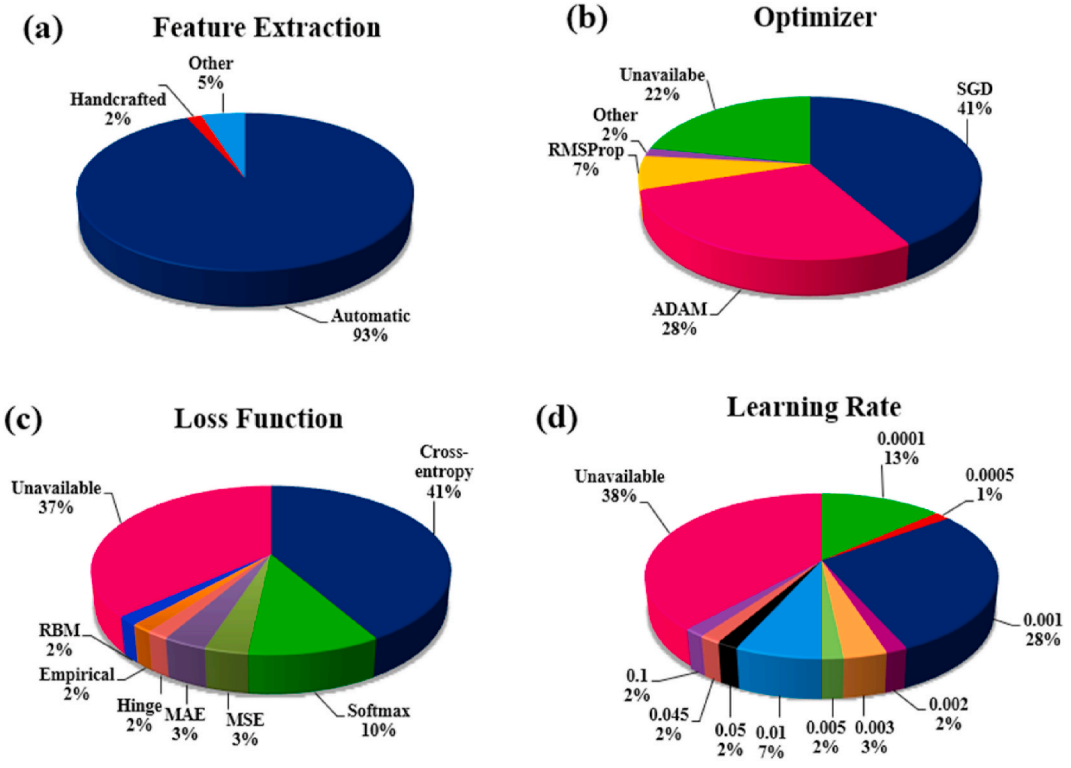

Fig. 13. HDL parameters: (a) Feature extraction, (b) Optimizer, (c) Loss function, (d) Learning rate.

probabilistic methods such as the Gaussian mixture model (GMM) and Kalman filter (KF) [35]. Generating these temporal feature frames can aid any temporal application (e.g., sport, HAR, and fall detection).

3.4.2. *Importance in HDL framework and types of optimizers*

The HDL models' performances are strongly influenced by the type of optimizer used. The right optimizer enables HDL models to work appropriately with our data by tweaking all the hyper parameter values. At the same time, it reduces the loss value and improves the performance of the model.

Fig. 13 (b) depicts the distribution of the various optimizers used in the HDL models in this study. The majority of the studies used stochastic gradient descent (SGD) (**41%**) [24,28,30,32,35,36,38,40,45,46,51,52,54,56,59,62,64,65,67,68,71,75,76,78], followed by adaptive moment (ADAM) (**28%**) [12,22,27,39,41,44,47,49,53,55,57,58,63,66,72,74,77], root mean squared (RMSprop) (**7%**) [13,26,37,73], and other (**2%**) [10].

Unfortunately, **22%** of the studies did not provide information about the optimizer used. This clearly explains the importance of optimization in HDL framework.

While SGD and ADAM are gradient-based optimizers, SGD works on a selected subset of the dataset (or a random selection of data) at a time rather than performing computations on a whole dataset—hence, it is much faster. Our observations showed that ADAM is the best choice among all optimizers, as it is the most widely applied to time-series and video data for temporal and spatial-temporal architectures [41,58].

3.4.3. Types of loss function in HDL framework

HDL models are intended to choose the right loss functions to fit the model and simultaneously minimize its value to obtain the optimized performance. The choice of the loss function in an HDL model depends on the activation function used in the output layer (Softmax or ML-based classifier) [21]. The majority of HDL models (across all classes) adapt cross-entropy (CE) and Softmax loss functions (variants of CE loss) since these HDL models are suitable for classification tasks (Fig. 13 (c).

Our observations show that the spatial HDL class [13,24,28,33,37,38,40,45,51,52,59,62,64,72] often uses cross-entropy loss (CE-loss), while temporal and spatial-temporal HDL classes use three kinds of loss functions: CE-loss [26,31,32,35,41,47,49,53,54,57,63,73–78], mean squared error (MSE) [22,39], and mean absolute error (MAE) [12,58]. The loss function complements the optimizer, which is controlled by stochastic gradient descent and minimizes the model's error. Thus, the loss function converges faster with optimization using gradient descent, thereby improving the model's performance [92]. At the end of each epoch during the training process, the loss will be calculated using the model's output predictions and the true labels for the respective input.

Another name for the CE-loss function is log-loss, and it is one of the most widely used loss functions. The value of CE-loss varies between 0 and 1, increasing when the predicted probability of the dataset starts to vary from the actual value. CE-loss can be measured mathematically as in Eq. (1):

$$CE-loss = -\sum_{c=1}^{M} y_{o,c} \log\left(p_{o,c}\right) \tag{1}$$

where M is the total number of classes, the log is the natural log, and y is the binary indicator (values vary from 0 to 1). The class label c is the correct classification for observation o, while p is the predicted probability for observation o, given the class c. Other loss function calculators that have been used in HDL models are MSE [22,39], MAE [12,58], and empirical [68], all of which are suitable for regression, prediction, and forecasting. In the MSE and MAE loss functions shown in Eq. (2) and Eq. (3) below, y_i and y_i^p are the target variable and predicted values, while N represents the number of samples.

$$MSE = \frac{\sum_{i=0}^{N}(y_i - y_i^p)^2}{N} \tag{2}$$

$$MAE = \frac{\sum_{i=0}^{N}|y_i - y_i^p|}{N} \tag{3}$$

The fourth type of loss is hinge loss (HL) [36]. This function, shown in Eq. (4), is closely associated with the SVM-based ML classifier.

$$HL = \max\left(0, 1 - y^i\left(x^i - \boldsymbol{b}\right)\right)\right) \tag{4}$$

In the above equation, y^i and x^i refer to the ith instance in the training set for the predicted and intended values of the classifier, while b is the bias term.

Empirical loss (EL), on the other hand, is calculated by averaging the loss function of a training dataset. It is represented mathematically as in Eq. (5), where $L(y_i, y_i^{\sim})$ is the loss function that measures the cost of predicting y_i when the actual answer is $y_i^{\sim}$. N is the number of samples.

$$EL = \frac{1}{N}\sum_{i=1}^{N} L\left(y_i, y_i^{\sim}\right). \tag{5}$$

3.4.4. Types of learning rate in HDL framework

The learning rate (LR) determines how far the neural network weights change within the context of optimization while minimizing the loss function. Thus, this parameter is important to optimizer and loss function. The commonly used initial LRs of various HDL models are listed in Fig. 13 (d). The most used suitable LR is 0.001 (adapted in **28%** of the studies) [13,22,27,31,33,36,41,46,47,53,68,72–77]. The other standard LR values are 0.0001 (**13%**) [12,26,28,37,38,55,57,62], and 0.01 used by (**7%**) [42,52,54,78]. The remaining LR values of the HDL are the minor variations of the above-said values.

The optimal LR for most of the HDL models is observed to be 0.001, as it provides suitable weights to optimize models' performance by reducing the error rate. A lower LR might allow the model to learn in a more optimal way, or even globally optimal sets of weights, but could also take significantly longer to train. So, it is essential to choose the LR carefully to suit the model's architecture. Note that for any input model (SDL or HDL), the number of layers, their organization, and other hyper parameters such as the number of epochs, batch size, and optimization are kept constant for any given LR.

4. Comparative analysis of the three HDL classes

An HDL model is classified according to the type of input modalities involved and the type of AI combination utilized to hybridize the model. A deeper analysis is required to compare the three HDL classes based on several attributes. Table 3 shows the characteristic distribution with respect to 14 attributes: (i) types of applications developed for specific HDL class, (ii) data types used for each class, (iii) types of imaging modalities used for that class, (iv) AI combination used to design the HDL, (v) stage one of the HDL models, (vi) stage two of HDL model, (viii) granular type (GT) used during classification, (ix) accuracy, (x) optimizer type (xii) loss function, (xiii) learning rate, and (xiv) the duration of the research study. This section will make a broad and deep analysis among the three HDL classes by considering these AI-based attributes.

4.1. Definition and characteristics for the formation of three HDL classes

Hybrid deep learning (HDL) which is categorically divided into three fundamental classes such as spatial, temporal, and spatial-temporal, is the mainstream focus in this narrative review. The formation of these three HDL classes and the AI architectures in these three classes are also interlinked with various other AI-based attributes. For a better understanding of the HDL literature in terms of the type of AI architecture, their performance, and categorization into three types of HDL paradigms, we first present **(a)** criteria adapted for the above classes and **(b)** AI-architectures used in each of these HDL classes.

(a) The role of Input Imagery

In the proposed review, two AI-based attributes are considered: (i) input imaging modality and (ii) AI-based architecture for categorizing the studies into the type of HDL class. These two attributes are considered because they provide a clean and independent solution for separating the HDL classes.

When the AI architecture accepts input as static imagery, such an HDL architecture can be categorized into a *spatial* HDL class. The input for the *temporal* HDL class is motion imagery and other forms of temporal data such as time-series data or video data. Similarly, the *spatial-temporal* class has both static imagery and motion imagery as an input modality.

(b) Relationship of architectural stages in HDL framework

Table 3
Characteristic distribution of three HDL classes based on various attributes.

HDL Class	Applications [Citation]	Data Type	IM	AI Combination	Stage-1 of HDL	Stage-2 of HDL	GT[a]	ACC[b]	OPT	LF	LR	RP
Spatial	Medical [10,11,13,23,28,37,38,45,52,55,62,64]	Image,	Natural,	DL + TL + ML,	LeNet,	Softmax,	Binary	92.9	SGD,	Cross-entropy	0.01,	2012–2021
	Face Reco [24,43,51,56].	Document,	CT,	DL + TL,	AlexNet,	SVM,	Multi		ADAM,		0.001,	
	Document Reco [50].	Files	PET,	DL + DL,	VGG,	KNN,			RMSProp		0.0001	
	Cyber Security [29,72]		MRI,	DL + ML	ResNet,	RF,			PSO-Guided			
	Agriculture [42,65]		X-ray		GoogleNet,	Ensemble			WOA			
	Multimedia [22,33,40,59]				MobileNet,	XGBoost						
					RBM							
Temporal	Time-Series Data [41,58]	Image,	Natural,	DL + DL,	RNN,	Softmax,	Multi	91.1	SGD,	Cross-entropy,	0.01,	2016–2021
	HAR [25,31,34,35]	Audio,	MRI,	DL + ML	LSTM,	LRL			ADAM,	MSE,	0.001,	
	HSI Analysis [69]	Video,	HIS		BiLSTM,				RMSProp	MAE	0.0001	
	NLP [54,77,78]	Text,			GRU,							
	Multimedia [49,63]	Document			RBM							
	IDS [32]											
Spatial-Temporal	Medical [70,74]	Image,	Natural,	DL + DL,	RNN,	Softmax,	Multi	89.5	SGD,	Cross-entropy,	0.01,	2016–2021
	HAR [12,26,44,57,60]	Audio,	MRI,	DL + TL + ML,	LSTM,	SVM,			ADAM,	MSE,	0.001,	
	HSI Analysis [27,30,47,48,53,71,73]	Video,	HIS	DL + ML	BiLSTM,	LRL			RMSProp	MAE.	0.0001	
	Traffic-Flow [39,66,67]	Text,			GRU,					Hinge,		
	Multimedia [36,46,68,75,76]	Document			RBM					Empirical		

[a]**GT**: Granular Type.
b**ACC** (%): Accuracy, **LRL**: Linear Regression Layer; **IDS**: Intrusion Detection System; **HAR**: Human Action Recognition; **NLP**: Natural Language Processing; **HIS**: Hyperspectral Image; **OPT**: Optimizer; **PSO**: Particle Swarm Optimization; **WOA**: Whale Optimization Algorithm; **IM**: Imaging Modalities; **LF**: Loss Function; **LR**: Learning Rate; **RP**: Research Period.

The purpose of an HDL architecture is to classify different applications based on the kind of input imagery used (such as spatial, temporal, or a combination of spatial-temporal). Such a classification process in the HDL framework requires input information that can be used to funnel into the classifier. Such information needs a refined set of signatures in the form of features. Thus, the HDL architecture must have at least two stages, stage-I and stage-II, where stage-I can be used to extract features and stage-II can be used for segregation or classification. Such a generalized paradigm can be tailored for three sets of HDL paradigms.

To assert this, we conclude that a *spatial* HDL paradigm will extract features from static imagery in stage-I while uses a classifier to classify the features in static mode in stage-II. Popular models which belong to this spatial paradigm are categorically and systematically uses a 2D convolutional neural network (CNN), while stage-II uses a fully connected layer (FCL) with softmax or a machine learning-based stand-alone classifier. Note that one can also adapt the transfer learning models as part of the stage-I, driven by the pre-trained weights, and tandemly connected to stage-II, which can be FCL with softmax or a machine learning-based stand-alone classifier. Popular examples of transfer learning models can be AlexNet [10,13,38,45,62,65], visual geometry group (VGG) [10,13,38,64], ResNet [10,28,38,43], or GoogleNet [10,22,28,38,59], which are well used in Deep Learning industry.

As the name reflects, the *temporal* HDL paradigm needs the architecture that can accept and process temporal information and deliver the results, ensuring the HDL system's ability to classify different kinds of semantics. There are architectures in Deep Learning that can effectively extract temporal features and high-level semantics from a sequence and discrete data. Such an intuition exists in a neural network under the process of recurrent paradigm. Thus *temporal* HDL class must also have stage-I and stage-II, where both can be of a recurrent type such as Long Short-Term Memory (LSTM) [32,49,54,58,63], Bi-directional LSTM (BiLSTM) [54,69,77,78], GRU, or bidirectional GRU (BiGRU) [31,35], and Simple Recurrent Unit (SRU) [31] and 3D-CNN [25,34]. However, it has been further noticed that architectures that are more computation efficient are restrictive in terms of the connection between the nodes that can be adapted as stage-I in a temporal framework. These networks have hidden layers whose activations can be used as an input to other models, thereby improving the overall performance. Such architecture follows the conventional Boltzmann Machines [25,41]. Thus, stage-I can be 3-D CNN or Deep Belief Networks (DBN), such as using the power of restrictive Boltzmann Machines (RBM).

This 3rd class of the three HDL classes adapts the fusion of *spatial and temporal* information for the end-product of classification. Due to the fusion activity, it is more complex compared to plain spatial and plain temporal paradigms. The fusion process still preserves the architecture of two stages, where stage-I plays a role in feature extraction while stage-II is used for classification. The key thing here to notice is that stage-I is no longer spatial in nature but uses both spatial and temporal jointly, while in stage-II, it further classifies the spatial domain and temporal domain jointly. Even though it is a fusion or joint process, but the stage-I and stage-II is still the accumulation of fundamental architectures, for example, in stage-I, spatial information can be processed using 2D CNN (such as AlexNet, VGG, ResNet, or GoogleNet), and temporal information can be processed using 3D CNN or RNN (such as LSTM, BiLSTM, GRU, and SRU). Stage-II, when connected to stage-I, supports classification and uses the same fundamental set architectures, for example, 3D CNN for classification of spatial features (such as AlexNet, VGG, ResNet, or GoogleNet), and classification of temporal information such as RNN.

4.2. Application-based HDL classes along with their AI-attributes: similarities and differences

Having defined the type of HDL classes by (a) input imaging modality and (b) AI architecture components in them, one can better appreciate when these two derivatives are linked to the type of application we choose.

Table 4
Example showing the different kinds of HDL applications in relation to other AI-attributes.

-	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
SN	**Applications Type**	**HDL Type**	**Image Modality**	**Stage-1 of HDL**	**Stage-2 of HDL**	**Hybridization**	**Optimizer Type**	**Loss Function**	**Learning Rate**	**Average Accuracy**
Example 4 (a)										
R1	Medical Application	Spatial [10,11,13,23,28,37,38,45,52,55,62,64]	X-ray, CT, MRI, Microscopy	CNN	CNN or ML	CNN + CNN, CNN + ML	SGD, ADAM, RMSProp	Cross-entropy	0.01, 0.001, 0.0001	94
		Spatial-Temporal [70,74]	EEG (MI), Histopathology	CNN	RNN (LSTM)	CNN + RNN (LSTM)				91.3
Example 4 (b)										
R2	Multimedia Application	Spatial [22,33,40,59]	Natural Image Type	CNN	CNN or ML	CNN + CNN, CNN + ML	SGD, ADAM, RMSProp	Cross-entropy	0.01, 0.001, 0.0001	93.38
		Temporal [49,54,63]	Video Type	3D-CNN, RNN	3D-CNN, RNN (LSTM)	3D-CNN+3D-CNN, SRU + GRU, GRU + GRU				98.4
		Spatial-Temporal [36,46,68,75,76]		2D-CNN, 3D-CNN, RNN	3D-CNN, RNN (LSTM)	2D-CNN + LSTM, 3D-CNN + LSTM,				77.5
Example 4 (c)										
R3	HAR Application	Temporal [25,31,34,35]	Video Type	3D-CNN, SRU, GRU	3D-CNN, SRU, GRU	3D-CNN+3D-CNN, SRU + GRU, GRU + GRU	SGD, ADAM, RMSProp	Cross-entropy, MSE, MAE	0.01, 0.001, 0.0001	98.59
		Spatial-Temporal [12,26,44,57,60]		2D-CNN, 3D-CNN	LSTM, BiLSTM	2D-CNN + LSTM, 3D-CNN + LSTM				94.46
Example 4 (d)										
R4	HSI Application	Temporal [69]	HS Image	LSTM	LSTM	LSTM + LSTM	SGD, ADAM, RMSProp	Cross-entropy	0.01, 0.001, 0.0001	85.73
		Spatial-Temporal [27,30,47,48,53,71,73]		2D-CNN, 3D-CNN	LSTM, BiLSTM	2D-CNN + LSTM, 3D-CNN + LSTM				98.68
Example 4 (e)										
R5	Document Recognition	Spatial [50]	Document	CNN	SVM	CNN + SVM	SGD, ADAM, RMSProp	Cross-entropy	0.01, 0.001, 0.0001	99.81
		Temporal [77,78]		BiLSTM	CNN	BiLSTM + CNN				84.1
R6	CyberSecurity	Spatial [29,72]	Image	RBM, CNN	SVM, RF	RBM + SVM, CNN + RF	SGD, ADAM, RMSProp	Cross-entropy	0.01, 0.001, 0.0001	99.98
		Temporal [32]	Time Series Image	CNN	LSTM	CNN + LSTM				98.43
Example 4 (f)										
R7	Time Series or Traffic Signal analysis	Temporal [41,58]	Time series and Video data	CNN, DBN	LSTM	CNN + LSTM, DBN + LSTM	SGD, ADAM, RMSProp	Cross-entropy, MSE, MAE	0.01, 0.001, 0.0001	97.5
		Spatial-Temporal [39,66,67]		CNN	LSTM	CNN + LSTM				Error Index values
Example 4 (g)										
R8	Face Recognition	Spatial [24,43,51,56]	Natural Image	CNN	SVM, DT, RBM	CNN + SVM, CNN + DT, CNN + RBM	SGD, ADAM, RMSProp	Cross-entropy	0.01, 0.001, 0.0001	95.28
R9	Agriculture	Spatial [42,65]	Natural Image	CNN	SVM, KNN	CNN + SVM, CNN + KNN	SGD, ADAM, RMSProp	Cross-entropy	0.01, 0.001, 0.0001	94

Thus, we have specially designed a table that can link the performance of these AI architectures considered in one of the HDL classes keeping the application in mind. In fact, the computer vision application and AI architectures go hand-in-hand for the three types of HDL classes (spatial, temporal, and spatial-temporal). We have therefore taken several real-world applications, as shown in Table 4. Note that for linking the performance of these AI architecture with several performance attributes of AI such as the type of the optimization method, the type of the loss function, the learning rate, and the overall accuracy. Thus, such a table is ideal for understanding the performance metric, AI-architecture used in a particular HDL class for the application-architecture pair. Hence, this table also offers the advantage of addressing the differences and similarities among the studies used in the HDL literature. Table 4 is elaborated in the form of the matrix, where the type of the application (column 1) utilizing different HDL classes (column 2) and linking with (i) architectures (column 4, 5, and 6), (ii) optimization paradigms (column 7, 8, and 9), and their (iii) performances (column 10). Interestingly, we observed that there were seven different kinds of applications, such as (i) medical, (ii) multimedia, (iii) human activity recognition (HAR), (iv) hyperspectral image (HSI) analysis, (v) document recognition, (vi) cybersecurity, and (vii) time-series data analysis, which could be used for demonstrating the linking process, while the two other applications such as (a) face recognition (b) agriculture are not considered in this process but still included as they are using a solo HDL class. By this process, we cover all applications used for analysis under this narrative review.

The *medical applications* [10,11,13,23,28,37,38,45,52,55,62,64,70,74] (row 1) are covered by spatial and spatial-temporal HDL classes, having differences in input modalities and the type of AI architecture. The *spatial* HDL class uses X-ray, CT, MRI, and Microscopy data utilizing mainly CNN for stage-I and CNN or ML for stage-II, respectively. In the *spatial-temporal* HDL class, EEG signals using motor imagery and histopathology image were used, using the CNN, RNN (LSTM) as stage-I and stage-II, respectively, when using the hybridization process. Thus, the hybridization process for the spatial HDL class was CNN + CNN or CNN + ML, while for the spatial-temporal HDL class, the hybridization used CNN + RNN (LSTM). Note that the performance between the two HDL paradigms shows the accuracies of 94% and 91% (column 10), even though both paradigms used cross-entropy-based loss function. One reason for low performance in the spatial-temporal HDL class is the noisy data during the EEG data collection [89,93,94]. Further note that due to the temporal nature of the EEG data sets, the architecture demands the LSTM paradigm (see Appendix Figure B2).

The second and most important application of HDL is *multimedia* [22,33,36,40,46,49,54,59,63,68,75,76] (row 2). We observed that this application was most popular and enveloped all three kinds of HDL classes. Only natural images were used as an imaging modality in the spatial paradigm; unlike in *temporal and spatial-temporal* classes, the studies used video imaging. The main architectures for these HDL classes are CNN-based [22,33,40,59], 3D-CNN/RNN-based [49,63], and 2D-CNN/3D-CNN/RNN-based [36,46,68,75,76] for stage-I, respectively, while CNN, 3D CNN/RNN, and 3D-CNN/RNN for stage-II, respectively. Note that spatial and temporal paradigms were relatively superior to spatial-temporal class (column 10), having the accuracies of 93.38%, 98.4%, and 77.5%. This can be attributed due to the architecture layout in different HDL classes. The reason for low accuracy for spatial-temporal for multimedia applications is also the architecture layout in which the information is lost when there is a fusion of both spatial and temporal features before classification [36,46,68,75,76].

The *third application* was in the *HAR* area [12,25,26,31,34,35,44,57,60] (row 3). This is one of the burning topics, where the input data is video-type. Typically, temporal information is adapted; however, the application has been seen using spatial combined with temporal information. The leading architecture used for both stage-I and stage-II for the *temporal* class includes 3D-CNN [25,34], SRU [31], and GRU [31,35] architectures. On the other hand, when using the *spatial-temporal* paradigm, stage-I was mainly 2D-CNN or 3D-CNN [12,26,44,60], and stage-II was LSTM or BiLSTM [12,26,44,57,60]. Due to apparent reasons for adapting SRU or GRU combined with SRU and GRU for both stages, the performance in the temporal domain has reached 98.59%, while in the spatial-temporal class, the accuracy was 94.46%. Note that even though the process was hybridized, the information from one imagery to another slightly affected the performance [12,26,44,57,60].

As shown in Table 4, the fourth application was in the field of *hyperspectral image analysis* [27,30,47,48,53,69,71,73] (HSI). The main HDL classes used for this application were temporal and spatial-temporal. For *temporal*, the architecture used was LSTM [69] for both stage-I and stage-II, and for *spatial-temporal* 2D CNN/3D CNN [27,30,47,48,53,71,73] and LSTM/BiLSTM [27,30,47,48,53,71,73] for stage-I and stage-II, respectively. These architectures showed the best performance for the *spatial-temporal* paradigm (98.68%), unlike the pure temporal paradigm (85.73%). Note that HSI is the only application that had very high performance in spatial-temporal combination. One possible reason is the role of entropy and color-matching functions, which are powerful attributes for information measurements [95].

The fifth application in the queue is *document recognition* [50,77,78], attempted by spatial and temporal HDL classes. Under spatial HDL the stage-1 and stage-2 of HDL are CNN and SVM [50], while under temporal HDL, the stages are BiLSTM [77,78] and CNN [77,78] as feature extractor and classifier, respectively. The performance of this application under these two classes, the spatial class provides better and high accuracy of 99.81%, while temporal accuracy is 84.1%. The low accuracy of document recognition applications under temporal HDL is because of the feature extraction of temporal images as compared to static images of spatial HDL.

The sixth application is *cybersecurity* [29,32,72]. The input under spatial HDL for this application is the natural image, and the time-series image is the input to temporal HDL [32]. The CNN and RBM are used for stage-1 [29,32,72], while ML-based classifiers such as SVM, RF [29,32,72] are used for stage-II. Similarly, a combination of CNN and LSTM was used as a hybrid network under temporal class to solve this application. In both cases of HDL, the results are very promising, with 99.98% for spatial and 98.43% for temporal. The RBM and SVM classifier for temporal and spatial are the rules for better performance.

The *time series and traffic signal analysis* [39,41,58,66,67] application used temporal data that the temporal and spatial-temporal HDL model can only solve. The approach and architecture of both HDL models were very similar. Under both HDL models, the combination of CNN and RNN (LSTM) was used as the architecture's components [39,41,58,66,67]. For performance analysis, the studies under the temporal domain used average accuracy with a value of 95.28%; however, the spatial-temporal studies used various error-index such as MSE and MAE [39,66,67].

The reaming broad categories of applications considered for the studies include *face recognition* [24,43,51,56] and *agriculture* [42,65], solely considering spatial HDL class. The input to these HDL models were natural images of the specific application. These HDL models used simple architecture connections like the DL-based CNN model for feature extraction and the DL-based classifier for the classification [24,42,43,51,56,65]. The average accuracy under these two categories of applications under spatial HDL class is the fair and high value of 95% and 94%, respectively [24,42,43,51,56,65]. The better performance of these applications is attributed to the architecture layout with respect to the input type.

4.3. The popularity and preference of HDL architecture hierarchy

The architecture of HDL is an integral component that decides the nature of the HDL class. As discussed earlier, the two components (stage-I: for feature extraction and stage-II: for classification) are the two vital components, and the hybridization of these two components forms the hybrid model that can solve numerous applications. The components also vary for different HDL classes of spatial, temporal, and spatial-temporal. The robustness and popularity of the hybrid model depend upon the components used in the hybrid process. On a note, here we

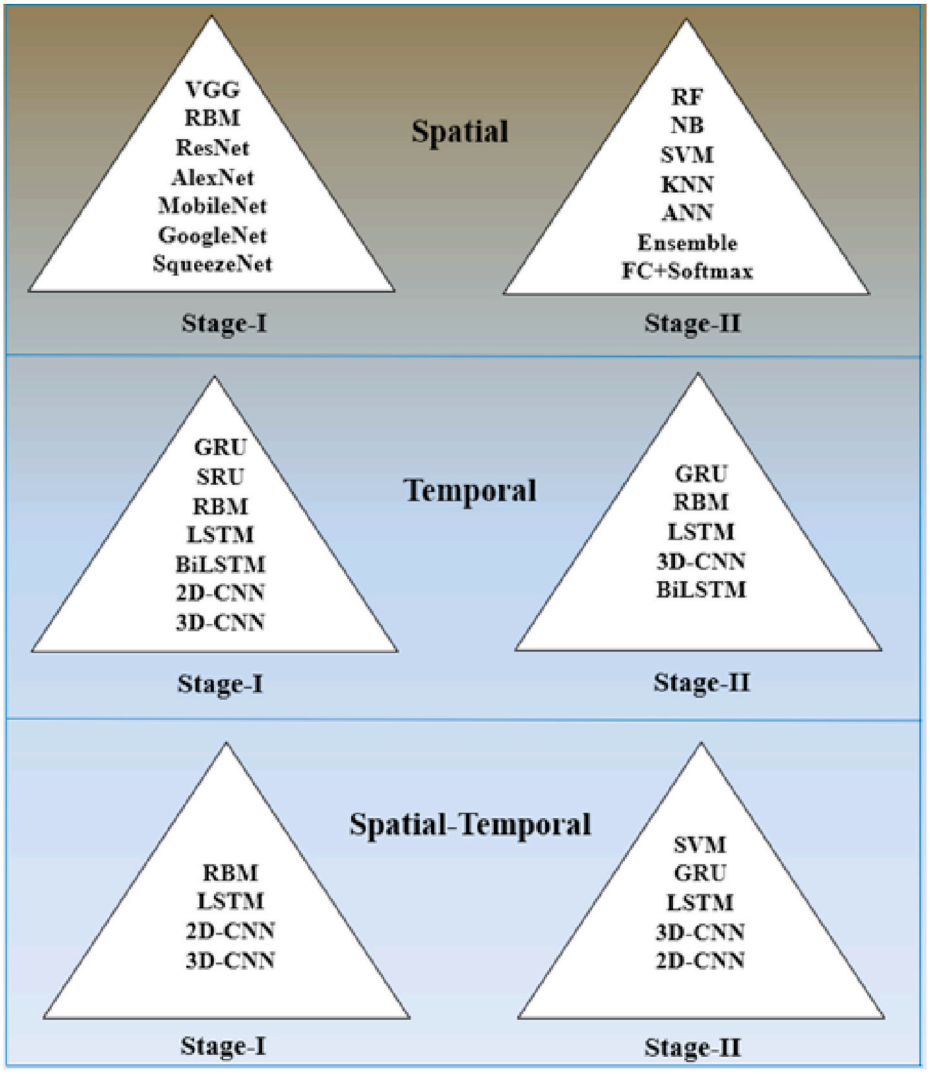

Fig. 14. Stage-I and stage-II hierarchical design based on the usage for various applications under various studies for spatial, temporal, and spatial-temporal HDL classes.

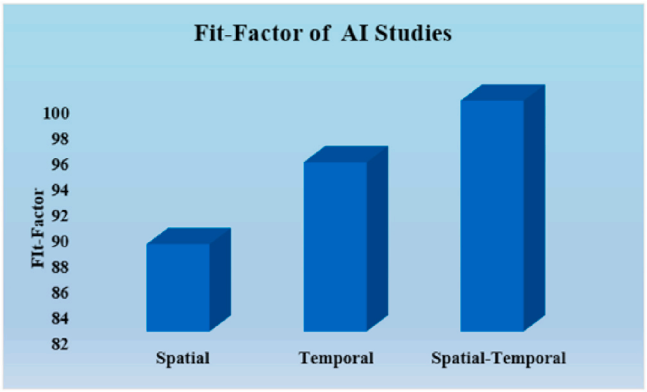

Fig. 15. Fit-Factor study of architectures used in HDL.

present the popularity of HDL components for all three HDL classes hierarchically based on the number of usages for various applications on different studies covered, as shown in Fig. 14.

4.4. The fit-factor study in three HDL classes

One of the key challenges is understanding how a particular study is selected and categorized into one of the three HDL classes. Even though one can choose the "application type" as a guiding force to refine the process of binning in three HDL classes, an alternative approach is based on an analytical solution such as designing a fitting paradigm to bin them in three HDL classes. This fitting paradigm can be formulated based on the generalized key AI attributes such as (a) input image modality, (b) type the application, (c) architecture type, (d) hybridization components of stage-I and stage-II, (e) data size. This fitting paradigm could also reflect the robustness of the study architecture to belong in the HDL class. Such fitting criteria can be formulated where the fit-factor

(FF) is a function of generalized key AI attributes for that study and can be symbolized as FF(s). Thus, each study can therefore compute the FF by grading the generalized key AI attributes on a scale of 0–1 and summed up. Thus, FF (s) value is computed for belonging to HDL class and can further be symbolized as FF(s,c), where s is the study into consideration and c is the class to which it will belong. The FF(s,c) is then computed for each study and summed as $\mathrm{FF(c)} = \sum_{s=1}^{S(c)} \mathrm{FF(s,c)}$, where S(c) is the total number of studies in each HDL class, c. The final sum is normalized to compute relative FF for each of the HDL classes (c). Using this analytical solution, the fit-factor values are shown in the bar chart below in Fig. 15. Note that the number of studies is different in various HDL classes. In our scenario, the S (spatial), S (temporal), and S (spatial-temporal) are 25, 13, 22, totaling 60 studies that participated in the FF study. Using the above formulation, the FF(c) is then converted to percentage values shown in Fig. 15, leading to 80.04%, 80.69%, and 90.18%, respectively, for spatial, temporal, and spatial-temporal HDL classes.

5. Performance of HDL models

The robustness of any HDL model can be reflected by its performance evaluation (PE) parameters. The standard PE parameters used by any HDL model for classification evaluation are accuracy, sensitivity, specificity, precision, F-1 score, the area under the curve (AUC), kappa statistics, and error-index. Apart from that, to check the robustness of the model, optimization paradigm evaluation is necessary. Again, the impact of hardware and software resources on the performance of the model has equal importance at the same time. Furthermore, above all, the risk-of-bias analysis of the study ushers the balanced use of various AI-based attributes used in that study.

5.1. Statistical distribution of performance evaluation (PE) parameters

This section presents the statistical distribution of PE for three independent HDL classes, which consist of finding the mean, median, largest, smallest, SD, and variance. Such behaviour helps in understanding the statistical distribution for eight different PE parameters: accuracy, AUC, specificity, sensitivity, precision, F-1, kappa, and error-index the three different HDL classes. Note that each HDL class has a different number of identified studies and 25, 13, and 22 for spatial, temporal, and spatial-temporal HDL classes. Accuracy is a very well-known parameter and is used by almost all DL and HDL models. Fig. 16 (a) shows the histogram distribution of PE parameters used in this study in decreasing order. Fig. 16 (b) depicts the accuracy parameter (mean ± sd) for the three classic HDL models.

Tables 5–8 show the statistical distribution *vs.* performance evaluation parameters for spatial, temporal, spatial-temporal, and complete HDL classes. In Table 5, the PE parameters of all 25 studies of spatial class [10,11,13,22–24,28,29,33,37,38,40,42,43,45,50–52,55,56,59,62,64,65,72] are considered. Their statistical distribution, including count, mean, median, largest, smallest, SD, and variances, are collected for all PE parameters. For example, the accuracy count of 21 indicates that 21 of 25 studies considered accuracy as a PE parameter. The mean, median, maximum, minimum, and SD of the accuracy of all 21 studies are given in column 1 (marked ACC). The statistical distribution of each of the other PE parameters is given similarly in subsequent columns. The error-index (EI) column remains blank, as there are uneven error indices in various studies. The mean accuracy of the spatial class is 92.9%, whereas the maximum and minimum accuracies are 100% and 73%, respectively.

Similar findings can be observed in Tables 6–8 for temporal, spatial-temporal, and complete HDL classes, including 13, 22, and 60 studies. Of these studies, 10, 16, and 47 assessed the accuracy and calculated mean accuracies of 91.1%, 89.5%, and 91.9%, respectively. Furthermore, the performance levels of some typical HDL architectures and their corresponding SLDs are provided in Appendix A to highlight the hypothesis that HDL is superior to SDL.

5.2. Behaviour of performance evaluation parameters in HDL classes

In the earlier section, we demonstrated the link between application and HDL classes while briefly discussing only the accuracy parameter (see Table 4), but did not highlight the other performance parameters (PE) independently in three HDL classes. The plots for these can be visualized from Figure C1 to Figure C6 (Appendix C). This ensures the readers how each HDL architecture is performing under a specific HDL class architecture. This is another way to compare and contrast all three types of HDL classes using performance parameters by analyzing all the studies simultaneously. This behaviour offers an exciting paradigm for knowing low- and high-performing studies (architectures).

5.2.1. Accuracy by HDL classes

We arranged the accuracies in increasing order for the studies present in spatial, temporal, and spatial-temporal paradigms, and this can be seen in Figure C1 (Appendix C). Accuracy of individual studies under *spatial* HDL class depicting an accuracy range from 73% to 100%. The lowest accuracy under the spatial HDL model is observed for Bharati et al., who hybridized VGG and spatial transformer network (STN) with CNN. The reason behind the low-performance is the use of a high-resolution image (1024 × 1024), and volumetric dataset of X-ray images on a poorly performing hybrid model was due to no pre-processing of the data (such as normalization) and no data augmentation protocol [23]. The high-performing studies with an accuracy of around 100% are due to standard spatial HDL architecture on moderate data amount and other good criteria such as data augmentation, pre-processing, hyper-parameter tuning, which are the key factors that improve the performances of the HDL model. The moderate performing

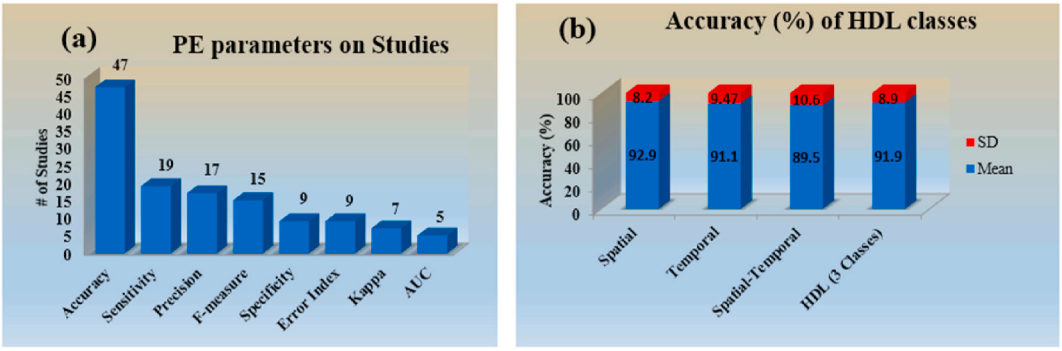

Fig. 16. (a) Histogram of PE parameters (b) Accuracy parameter for three HDL classes.

Table 5
PE metrics for the *spatial* HDL class only.

	ACC	SPE	SEN	PREC	F-1	AUC	KAPPA	EI
Count	21	9	11	9	7	4	1	3
Mean	92.9	97.8	92.9	90.5	92.73	0.98	100	–
Median	94.66	98.73	95	90.95	94.4	0.98	100	–
Largest	100	100	100	100	100	1	100	–
Smallest	73	92.92	63	69	68	0.97	100	–
SD	8.2	2.76	11.2	9.59	8.89	0.01	0	–
Variance	67.28	7.64	125	91.97	80.64	0	0	–

*SD: Standard Deviation, ACC: Accuracy, SPE: Specificity, SEN: Sensitivity, PREC:Precision, El: Error Index.

Table 6
PE metrics for the *temporal* HDL class only.

	ACC	SPE	SEN	PREC	F-1	AUC	KAPPA	EI
Count	10	0	5	4	6	0	1	2
Mean	91.05	–	97.3	92.24	91.86	–	72	–
Median	94.53	–	98.2	91.5	92.33	–	72	–
Largest	99.8	–	100	100	100	–	72	–
Smallest	82.14	–	82	83	80	–	72	–
SD	9.47	–	12	6.6	7.5	–	0	–
Variance	89.81	–	144	43.62	56.37	–	0	–

*SD: Standard Deviation, ACC: Accuracy, SPE: Specificity, SEN: Sensitivity, PREC:Precision, El: Error Index.

Table 7
PE metrics for the *spatial–temporal* HDL class only.

	ACC	SPE	SEN	PREC	F-1	AUC	KAPPA	EI
Count	16	0	3	4	2	1	5	4
Mean	89.48	–	90.7	90.5	88.83	0.89	94.65	–
Median	93	–	92	90.65	91.08	0.89	98.87	–
Largest	99.9	–	93	99	92.5	0.89	99.89	–
Smallest	71.43	–	87.2	81.7	90	0.89	81	–
SD	10.62	–	3.12	7.22	5.56	0	9.14	–
Variance	112.9	–	9.79	52.19	30.99	0	83.63	–

*SD: Standard Deviation, ACC: Accuracy, SPE: Specificity, SEN: Sensitivity, PREC:Precision, El: Error Index.

Table 8
PE metrics for *all* HDL classes.

	ACC	SPE	SEN	PREC	F-1	AUC	KAPPA	EI
Count	47	9	19	17	15	5	7	9
Mean	91.9	97.8	93.4	91.58	92.77	0.96	91.77	–
Median	95.31	98.73	95	92.2	96.1	0.98	98.87	–
Largest	100	100	100	100	100	1	100	–
Smallest	71.43	92.92	63	69	68	0.89	72	–
SD	8.9	2.76	9.46	8.35	9.31	0.04	12.18	–
Variance	79.21	7.64	84.3	69.85	86.82	0	128.89	–

*SD: Standard Deviation, ACC: Accuracy, SPE: Specificity, SEN: Sensitivity, PREC:Precision, El: Error Index.

studies and their accuracies are given in Figure C1 (a) (Appendix C), which uses the combination of architectures between 2D-CNN and ML-based classifiers on various static imagery data sets. Accuracy of individual studies under *temporal* class depicting an accuracy range from 82.14% to 99.9%, as shown in Figure C1 (b) (Appendix C). We can observe that all of the studies resulted in a fair accuracy from their corresponding temporal HDL architecture. The accuracy of individual studies under *spatial-temporal* class depicting values that range from 71.43% to 99.9%. Even if most of the studies perform well under spatial-temporal HDL architecture, few studies have accuracy below 80%, which may be a matter of concern. This can be attributed to the architecture layout in the HDL class, in which the information is lost when there is a fusion of both spatial and temporal features before classification. These low accuracies are observed for multimedia applications of audio-visual emotion detection and facial expression recognition.

5.2.2. *Specificity by HDL classes*

This performance parameter is found to be used only for *spatial* HDL class. However, we observed no specific reason for not being recognized by the other two HDL classes. Coincidentally, none of the studies under *temporal* and *spatial-temporal* class are using specificity. Note that the values generated by this parameter in *spatial* class are relatively higher, ranging from 92.92% to 100%, having an average of 97.8%, as depicted in Figure C2 (Appendix C) and Table 5 also.

5.2.3. Sensitivity by HDL classes

All the studies under spatial HDL class had fair and high values except one. We already discussed that study under accuracy, for which that study also had low values. The same reason can be applicable here also for sensitivity. Sensitivity values of individual studies under temporal HDL class depict scores ranging from 82 to 100%, as shown in Figure C3 (Appendix C). Only five studies under the HDL class have considered sensitivity. It is observed that accuracy is the standard classification performance measurement parameter used by almost all studies. Along with accuracy, other parameters are considered by studies that are good studies showing their robustness. We can observe that all of the studies resulted in fair sensitivity values under the temporal HDL class. Again, only a few studies use sensitivity as performance parameters under spatial-temporal HDL class, and all the three studies have a fair score, and there are no significant differences among these values.

5.2.4. Precision by HDL classes

All studies under spatial HDL class have got fair and high values except one. We already discussed that single study under the parameters: accuracy and sensitivity, for which that study also got a low score. The same reason can be applicable here also for precision. Precision values of individual studies under temporal HDL class depict scores ranging from 83 to 100%, as shown in Figure C4 (Appendix C). We can observe that all four studies resulted in fair precision values under the temporal HDL class. Again, only four studies used precision as performance parameters under the spatial-temporal HDL class, and all four studies have a fair score ranging from 81.7 to 99%.

5.2.5. F1-score by HDL classes

The F1-score for all the studies under spatial HDL class has got fair and high values except one. We already discussed that study under accuracy sensitivity and precision, for which that study also got low values. The same reason can be applicable here also for the F1-score. F1-score values of individual studies under temporal HDL class depict scores ranging from 80 to 100%, as shown in Figure C5 (Appendix C). We can observe that all of the studies resulted in fair F1-score values under the temporal HDL class. Again, only a few studies use F1-score as performance parameters under spatial-temporal HDL class, and all three studies have a fair score.

5.2.6. AUC, kappa statistics and error-index by HDL classes

The parameter AUC is observed to be used by only the spatial HDL class. Also, one study under spatial-temporal used the AUC as the parameters along with other performance parameters. Here, the AUC values observed from the studies are fair and high, as anticipated, as shown in Figure C6 (a) (Appendix C). Similarly, Kappa is used as a performance metric by spatial-temporal and temporal HDL classes. The values observed under this metric for the spatial-temporal class are fair and high ranging between 81 and 99.89%, as shown in Figure C6 (b) (Appendix C). The error-index (EI) column (Tables 5–8) remains blank as there are uneven error indices such as mean square error, mean absolute error, regression error used in various studies.

5.3. HDL optimization and risk-of-bias

5.3.1. Optimization paradigm: optimizers, loss functions, and learning rates

It is important to note that there are similarities between the various HDL applications, main attributed due to (a) optimizers (b) loss functions (c) learning rate. As discussed in section 3.4 regarding the importance of optimization algorithms in HDL framework, we simply highlight here in relation to risk-of-bias. The group of standard and common optimizers among the three HDL classes is SGD, ADAM, and RMSProp. However, gradient-based optimizers such as ADAM and SGD were suitable across all HDL classes. Regarding the loss function, cross-entropy was the first-choice of loss function within HDL classes for all kinds of applications. The MSE, MAE, is the other set of loss functions that also has been used for all HDL classes depending upon the application selection. Similarly, nearly all applications using HDL classes showed the initial learning rates as 0.01, 0.001, and 0.0001. One reason for these optimization paradigms is due to the standardization adapted in the HDL industry.

5.3.2. Risk-of-bias in HDL models

Risk-of-bias assessment is a popular parameter used to calculate the overall fitness of AI-based models. It is based on the biased nature of any study (e.g., related to the features of the study design or the way it is conducted), which can yield misleading results. So, this assessment explains that people should consider all the model parameters with equal importance while evaluating a model. Doing this ensures that the results are neutral and fair, thus helping the model become widely accepted [96,97]. We checked the risk-of-bias of all the reviewed studies by gathering important AI attributes regarding the regularization of the HDL model's methods to avoid over-fitting and then induced performances. This parameter includes batch normalization (BN), dropout, and early stopping. The statistics show that 90% of the models use dropout as the common regularization method, with some mentioning the use of BN [12,13,26,28,34,35,39,40,45,54,55,57,59,64,66,68,77,78]. A few others utilized the early stopping criterion [13,28,45].

In the next phase, we consider the hyperparameters of HDL models such as optimizer, learning rate, loss function, batch size, epochs, and average training time. SGD is the most desirable optimizer use, as it was employed in **41%** of reviewed studies [24,28,30,32,35,36,38,40,45,46,51,52,54,56,59,62,64,65,67,68,71,75,76,78], followed by ADAM (**28%**) [12,22,27,39,41,44,47,49,53,55,57,58,63,66,72,74,77]. The most common LR value is 0.001 (**28%**) [13,22,27,31,33,36,41,46,47,53,68,72–77]. The other commonly used LR values are 0.0001 (**13%**) [12,26,28,37,38,55,57,62] and 0.01 (**7%**) [42,52,54,78].

The standard loss function used across all the HDL models is cross-entropy (CE). However, MSE, MAE, and hinge loss are other options. The standard mini-batch sizes deemed suitable for HDL models are 32 and 64. While the epochs value ranges from 10 to 1000, it has been determined that using about 100 epochs is desirable. In general, the reviewed studies provide inconsistent information about the use of factors essential to evaluating a model, such as learning rate, loss function, batch size, epochs, and training time in the study.

In another phase of attributes, studies considered benchmarking [10–13,22–24,26,27,29–33,35–42,44,45,47–59,63–78], validation based on other datasets [12,24,27,29,30,32,35–38,43,44,47–49,51,53,56–58,65,68,71–73,77,78], clinical validation (for medical applications) [38], usage of statistical studies [10,12,32,39,49,54,63], and the number and size of datasets used.

The next set of attributes includes the HDL components, including the activation function, pooling function, number of convolution layers, and fully connected layers. In nearly all studies, the activation function was ReLU, and the standard pooling function was max-pooling. There is no specific rule for determining the number of layers, convolution layers, and fully connected layers to use in an HDL model. The goal is to minimize the number of layers to reduce the complexity of the model while enhancing its performance.

Finally, the most critical parameters used when evaluating an HDL model are PE parameters (e.g., accuracy, precision, and recall) and

performance analysis parameters (e.g., confusion matrix (CM) and ROC curve. The number of PE parameters included in a study to ensure the correctness of the model indicates the model's robustness. We determined that 47 out of 60 studies used accuracy as the standard parameter to check the correctness of models. We also observed that 90% of studies used CM to analyze their models' performances.

Based on the risk-of-bias analysis, following studies were acceptable in three classes: spatial [10,11,13,38,62,64,65], temporal [31,35,54,58,63], and spatial-temporal [36,46,47,68,75,76]. Thus, it can be concluded that some of the studies are biased because they did not consider all the critical parameters when evaluating their HDL models. Here, it is suggested that authors consider all the essential parameters to ensure the fair evolution of their models. More detailed analysis is required regarding risk-of-bias models.

5.4. Hardware and software considerations in HDL models

Hardware and software resources are essential factors linked to HDL models. As we reviewed all the studies, we found some that used a single dedicated NVIDIA GPU-based server along with dedicated CPU computational power [22–24,26,28,30,33,37,40,44–46,48,52,54,55,57,59,60,62,63,66–68,70–76]. However, some studies with more complex HDL stages and big datasets [26,36,51,55,65] used more than one dedicated GPU along with a CPU computation machine. Basically, temporal and spatial-temporal HDL classes, which are based on applications like HSI analysis, time-series data, audio-visual data, video surveillance data, and HAR, also utilize a dedicated workstation or server machine with a powerful GPU computation machine in combination with a CPU machine with high computational power [26,36,46,53,68,75,76].

Moreover, some of the studies on spatial HDL class used CPU-based devices with extra memory power and processor speed [29,38,42,43]. Also, studies belonging to any HDL class with a small amount of data took advantage of robust CPU-based computations [29,31,32,35,38,42,43,58,63,69]. Encouragingly, one study [49] used cloud computing platforms like Google Colab. However, several studies did not mention the hardware resource used [10–13,25,27,34,39,41,50,56,77,78].

Note that the studies adapted high-level languages like Python or MATLAB. These are considered standardized software tools and, thus, are unlikely to be biased. It is observed that more studies used Python [12,23,30,31,33,37,40,47,49,51,52,54,57–60,63,64,66,70,72–74] than MATLAB [13,29,35,38,42,43,45,55,62,65,69] or the convolutional architecture for fast feature embedding (CAFFE) toolbox [24,36,68,76]. The standard neural network framework is tensor flow (TF) with Keras software [23,49,54,58,63,64]. Some studies did not mention the software resource used [10,11,25–28,32,34,39,41,44,48,50,53,56,67,71,75,77,78].

6. Discussion

This study presented the first narrative review of its kind on HDL. Using the PRISMA model (Fig. 2), we were able to identify publications from its inception in 2012 to today. The primary objective of this review was to understand the role of fusing two different types of AI paradigms (e.g., spatial, temporal, and spatial-temporal) to meet three different core applications (Fig. 3). Several applications have evolved in HDL models, gradually expanding their usage, architecture, and hardware, thus improving their performance.

We demonstrated 14 different AI attributes in each of the three fundamental HDL classes demonstrating nearly 20 applications (Table 3). The reviewed studies were mainly published by IEEE and Elsevier in the Asian continent (Fig. 3). We also observed that the distribution pattern of each of the three HDL classes was uniform in terms of data size. We showed that HDL models had been dominated by automated feature extraction, ADAM optimizer, CE-loss function, and a learning rate of 0.001 (Fig. 15). Lastly, though our hypothesis was based on the core links between multiple stages in HDL models, we believe that HDL paradigms are superior to SDL paradigms based solely on the type of application chosen (Table 1). We anticipate this field to evolve in the coming decade.

6.1. Benchmarking

There are not many review articles in the area of HDL. One reason for this is that different groups from different places are involved in developing SDL models, and they design models at their own pace and for unique applications. However, research in the HDL area expanded rapidly in 2020, and we anticipate more activity in terms of review articles soon.

The second reason is that HDL model development is focused on specific data sets. This was clearly and recently evident by the first HDL review paper by Alzahab et al. [16], where the authors focused on a specialized area of the brain-computer interface (BCI). Due to its narrow focus, the HDL coverage was not generalized enough to cover a full review. The authors did not incorporate an exhaustive study on HDL, and their coverage of different applications was limited. Thus, their narration was too specific. Due to this limitation, they missed several HDL articles, and their review was limited to 47 studies. Note that since Alzahab et al. covered only the BCI area, EEG was considered only under the umbrella of healthcare applications.

Differently, our study covered a large number of healthcare applications such as COVID-19, brain cancer, EEG, breast cancer, lung cancer, and diabetes. Since Alzahab et al.'s application is specific to BCI, their focus was on temporal HDL architectures and a few spatial-temporal HDL architectures. However, the present review focused on three fundamental classes of HDL models (i.e., spatial, temporal, and spatial-temporal). Due to the limitations of Alzahab et al.'s study, the main architectures used in that study were CNN, RNN, and DBN. Meanwhile, we considered a wide range of hybridization paradigms, such as CNN (2D-CNN, 3D-CNN), SDL (AlexNet, VGG, ResNet, GoogleNet), RNN (LSTM, BiLSTM, SRU, GRU), and DBN. Lastly, Alzahab et al. presented limited performance parameters for evaluating HDL models, whereas we used eight different PE parameters.

6.2. A special note HDL classes and applications

HDL has provided new opportunities in the field of AI. It has provided immense flexibility for designing various applications due to its inherent ability to bridge different architectures. Because HDL models can use either cascading different architectures or tandem connections between SDL and ML-based classifiers, the better of the two can be used in any given case to improve the model's performance.

The power of feature extraction using SDL when bridged by ML classifiers straightforwardly and elegantly boosts the performance of HDL models. One of the most prominent aspects of HDL is the zig-zag puzzle nature of the connections, which provides a large and diverse number of applications, such as hyperspectral image classification, audio-visual emotion recognition, human activity recognition, time-series data analysis, traffic flow analysis, video surveillance, and

healthcare applications. As the number of applications grows, we expect upgraded hardware to accommodate new HDL architecture designs. In fact, this could be a reversible solution by which HDL will grow further into a cascade of DLs to fit the miniature hardware like a cascaded set of filters to yield true signal recovery from noisy signal input data. The three fundamental HDL classes are likely to evolve from their current linear nature to an exponential nature as multi-disciplinary mergers increase the robustness of applications' output.

6.3. Open-ended challenges, an extension of HDL, and recommendations

Imaging science has been applied to the COVID-19 pandemic and other global health challenges. However, it has several important challenges in itself, such as the incomplete understanding of how HDL reflects the severity of COVID-19 in CT lung scans and—more specifically—the locations (like global positioning system) of COVID-19 severity in 2D and 3D lung scans [98].

It would be interesting to explore whether HDL can predict future symptoms for heart disease [99], cardiovascular risk [100], heart rate variability [101], plaque tissue characterization [102,103], and stroke using low-cost ultrasound methods [104,105], which is the number one killer on this globe. HDL could also be applied diabetes [106]; human blindness [107]; and thyroid [108], liver [109,110], prostate [111,112], ovarian [113–116], lung [117], brain [118], breast [119,120], and skin cancers [121] in developing countries. Moreover, Alzheimer's disease [122], multiple sclerosis [123], and Parkinson's disease represent other health issues that [124] have not been addressed using HDL and, thus, require exploration. Clearly, a gamut of HDL healthcare applications is waiting to be studied.

AI performance attributes, such as speed, accuracy, hardware miniaturization, performance, ensuring automated designs, efficacy, and safety, are of prime importance during the commercialization of HDL technologies [112], which are still in their infancy [125]. Thus, the trade-offs between these AI performance-based attributes still need to be shown.

6.4. Strengths, weakness, and extensions

The main strength of this study is that it selected a valuable set of **127** articles that aid our understanding of the HDL paradigm, particularly in comparison with SDL. The present study provided insights into the architectural differences between the three HDL classes based on 14 types of AI attributes. Also, the present study was has enhanced our understanding of nearly 20 different HDL applications considering the three classes of interest.

Even though the study had a positive outcome, the benchmarking section could not be elaborated due to the limited number of HDL reviews. Further, the HDL reviews discussed here did not highlight the risk-of-bias component, which is an integral part of AI paradigms. However, we anticipate more systematic reviews using HDL concepts to be published in the near future. Finally, we expect HDL applications to become increasingly diverse as computer vision techniques evolve, leading to, for example, larger data frameworks [19], cloud-based techniques [126], faster hardware, and miniaturized technologies.

7. Conclusion

HDL is the innermost component of the AI cycle's evolution (e.g., machine learning, deep learning) and is influenced by the ability of any given application. We designed HDL models of three different classes (i.e., spatial, temporal, and spatial-temporal) based on evaluations in imaging science. Exhaustive statistical data were displayed as pie charts and bar charts from supporting inferences regarding numerous variables, including the type of application, publisher, continent, optimizers used, hyperparameters adapted, and data size. Finally, we demonstrated that HDL models outperform SDL models. However, more evidence needs to be gathered to confirm this finding. As a final note, we expect the application of HDL models to grow and change significantly, and perhaps they will eventually behave semi-exponentially.

Disclosure

Dr. Jasjit S. Suri is affiliated with AtheroPoint, Roseville, CA, USA, dedicated to Stroke and Cardiovascular Imaging.

Declaration of competing interest

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.

Appendix A

Table A.1
SDL *vs.* HDL performance.

SN	Author	Dataset	SDL Model	SDL ACC*	HDL Model	HDL ACC*
1	Rezaee et al. [13]	COVID-19 CXR	Inception	95.3	CaffeNet,	99.5
			DenseNet201	95.3	Alexnet,	
			ResNet101	94.1	VGG-f + SVM	
			Vgg16	97.9		
2	Elkorany et al. [11]	COVID-19	SqueezeNet	89.45	COVIDetection	94.45
		Chest CT	ShuffleNet	91.95	Net	
3	EL-Kenawy et al. [10]	COVID-19	AlexNet	79	SFS-Guided	0.995[&]
		Chest CT	VGG16Net	58.21	WOA	
			VGG19Net	77.17		
			GoogleNet	73.06		
			ResNet-50	77.17		
4	Vijayalakshmi et al. [64]	MAMIC database	VGG16	87.84	VGG16 + SVM	89.21
			VGG19	91.47	VGG19 + SVM	93.13
5	Togacar et al. [62]	Cancer Imaging	AlexNet	89.14	AlexNet + kNN	95.51
Mean				**85.50**		**95.21**
±SD				**±11.53**		**±3.94**

*Accuracy (%).
&AUC, +: Hybrid, SFS: Stochastic Fractal Search, WOA: Whale Optimization Algo., VGG: Visual Geometric Group, *MIMIC*: Mortality in *Malaria* Intensive Care, COVID: Carona Virus Disease, CXR: Chest X-ray.

Appendix B

Artificial neural network

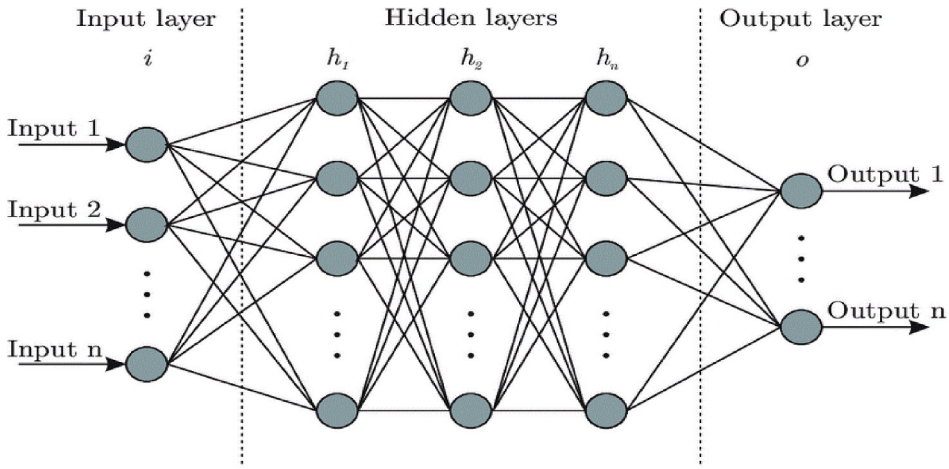

Figure B1. Basic flow diagram of an artificial neural network mimicking the human brain of the bio logical neural network [127].

Long short-term memory networks

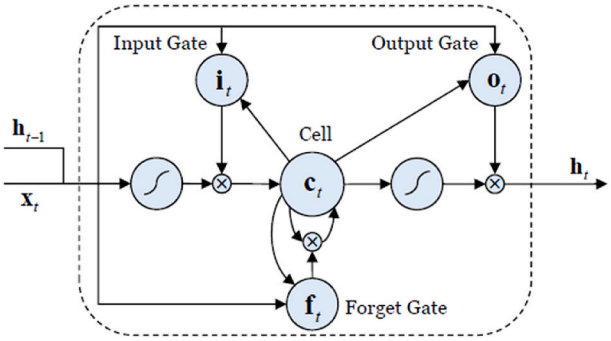

Figure B2. Structure of an LSTM unit that brings the core idea of the built-in memory cell, which is an integral component of temporal or spatial-temporal architecture [68].

Gated recurrent unit

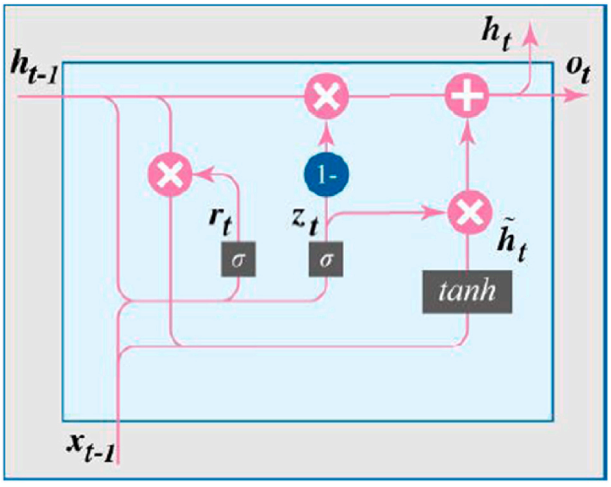

Figure B3. A typical structure of a GRU with memory serving as an integral component of temporal or spatial-temporal architecture, as it solves the vanishing gradient problem [31].

Restricted Boltzmann machine

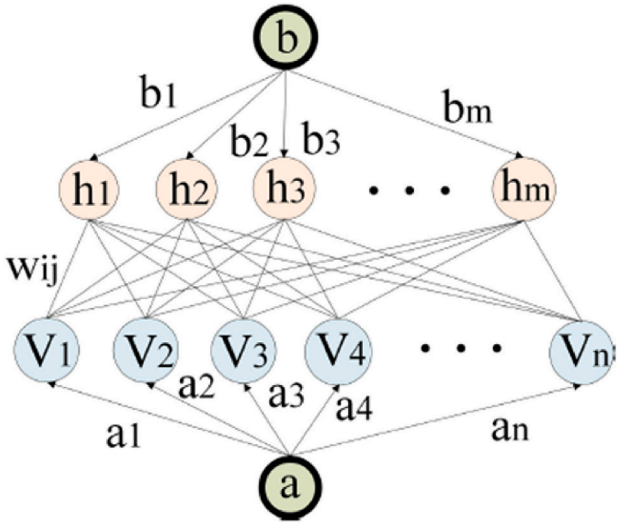

Figure B4. The RBM structure of a deep belief network used as components in some special HDL architectures [41].

Appendix C

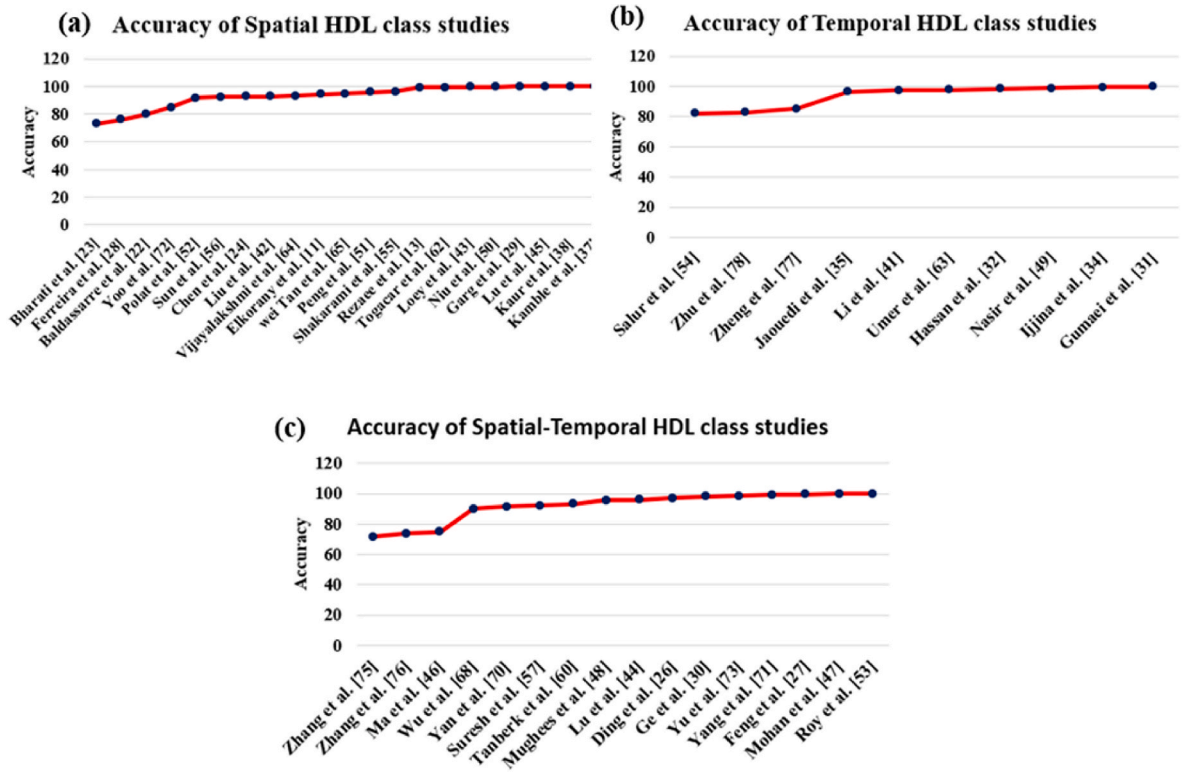

Figure C1. Spatial, Temporal, and Spatial-Temporal distribution of performance metric – Accuracy. Top: Spatial, Temporal; Bottom: Spatial-Temporal. Accuracies are arranged in ascending order for each of the HDL classes.

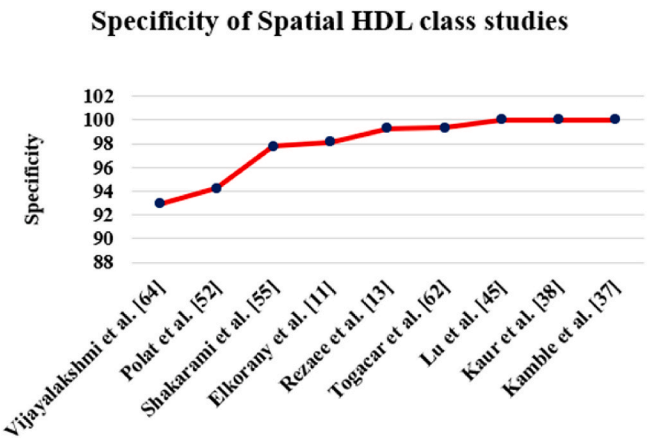

Figure C2. Various studies using specificity as performance parameters under spatial HDL class.

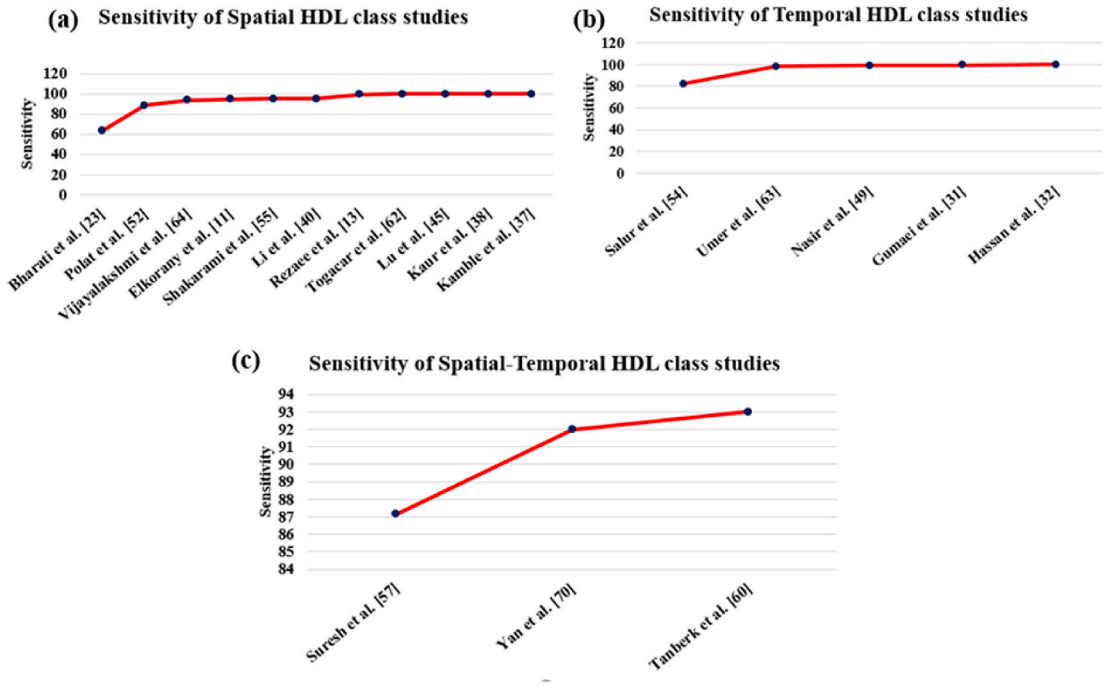

Figure C3. Spatial, Temporal, and Spatial-Temporal distribution of performance metric – Sensitivity. Top: Spatial, Temporal; Bottom: Spatial-Temporal. Sensitivity is arranged in ascending order for each of the HDL classes.

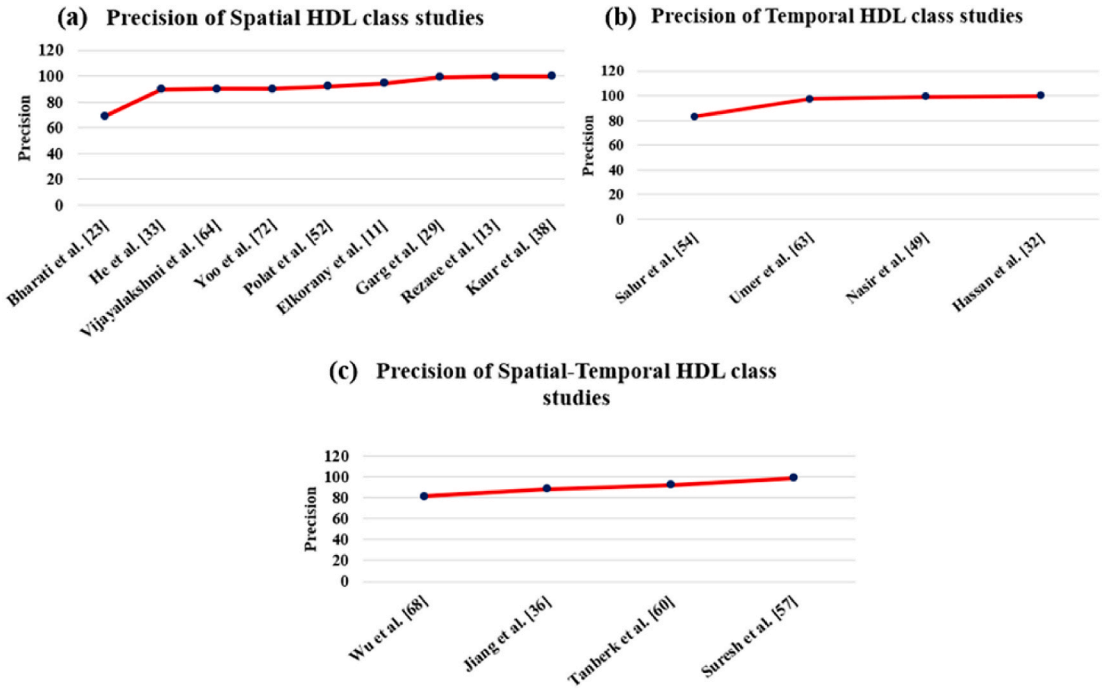

Figure C4. Spatial, Temporal, and Spatial-Temporal distribution of performance metric – Precision. Top: Spatial, Temporal; Bottom: Spatial-Temporal. Precision is arranged in ascending order for each of the HDL classes.

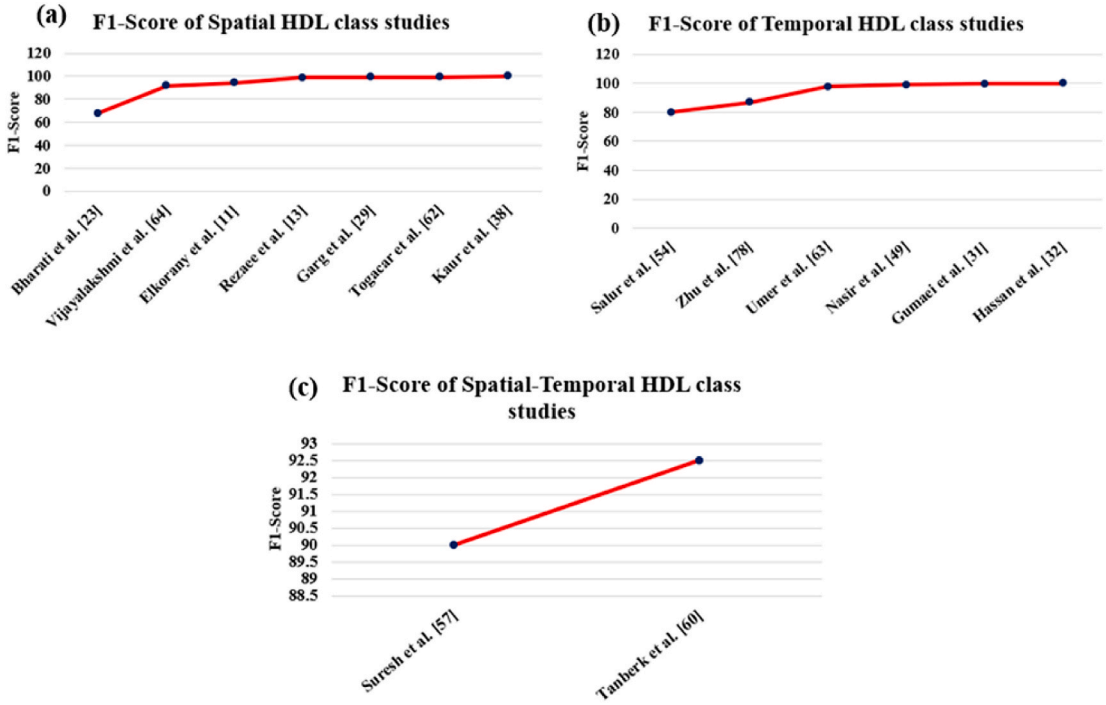

Figure C5. Spatial, Temporal, and Spatial-Temporal distribution of performance metric – F1-Score. Top: Spatial, Temporal; Bottom: Spatial-Temporal. F1-Score is arranged in ascending order for each of the HDL classes.

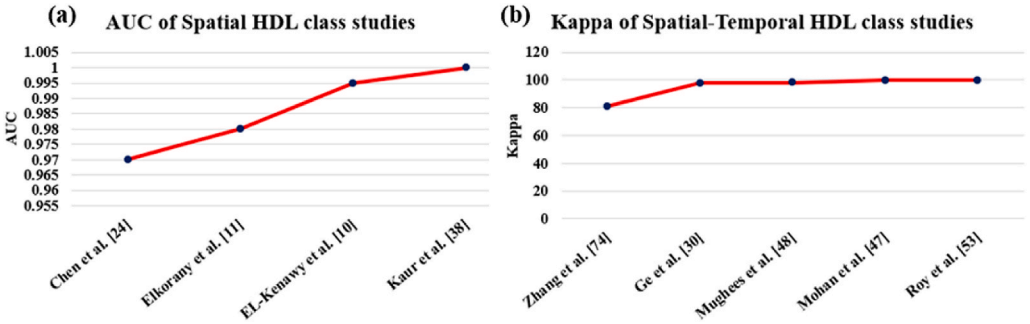

Figure C6. AUC (a) and Kappa (b) values for studies under spatial and spatial-temporal HDL class.

References

[1] N. Haefner, J. Wincent, V. Parida, O. Gassmann, Artificial intelligence and innovation management: a review, framework, and research agenda☆, Technol. Forecast. Soc. Change 162 (2021) 120392.

[2] C.C. Bennett, K. Hauser, Artificial intelligence framework for simulating clinical decision-making: a Markov decision process approach, Artif. Intell. Med. 57 (1) (2013) 9–19.

[3] M.H. Jarrahi, Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making, Bus. Horiz. 61 (4) (2018) 577–586.

[4] L.A. Lynn, Artificial intelligence systems for complex decision-making in acute care medicine: a review, Patient Saf. Surg. 13 (1) (2019) 1–8.

[5] G. Marcus, E. Davis, A.I. Rebooting, Building Artificial Intelligence We Can Trust, Vintage, 2019.

[6] J.L. Duffany, Artificial intelligence in GPS navigation systems, in: 2010 2nd International Conference on Software Technology and Engineering, vol. 1, IEEE, 2010, pp. V1-V382–V1-V387.

[7] M. Schedl, Deep learning in music recommendation systems, Frontiers in Applied Mathematics Statistics 5 (2019) 44.

[8] H. Khayyam, B. Javadi, M. Jalili, R.N. Jazar, Artificial intelligence and internet of things for autonomous vehicles, in: Nonlinear Approaches in Engineering Applications, Springer, 2020, pp. 39–68.

[9] P. Gentsch, Conversational ai: how (chat) bots will reshape the digital experience, in: AI in Marketing, Sales and Service, Springer, 2019, pp. 81–125.

[10] E.-S.M. El-Kenawy, A. Ibrahim, S. Mirjalili, M.M. Eid, S.E. Hussein, Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images, IEEE Access 8 (2020) 179317–179335.

[11] A.S. Elkorany, Z.F. Elsharkawy, COVIDetection-Net: a tailored COVID-19 detection from chest radiography images using deep learning, Optik 231 (2021) 166405.

[12] Y. Karadayi, M.N. Aydin, A.S. Öğrencí, Unsupervised anomaly detection in multivariate spatio-temporal data using deep learning: early detection of COVID-19 outbreak in Italy, IEEE Access 8 (2020) 164155–164177.

[13] K. Rezaee, A. Badiei, S. Meshgini, A hybrid deep transfer learning based approach for COVID-19 classification in chest X-ray images, in: 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME), IEEE, 2020, pp. 234–241.

[14] M. Biswas, et al., State-of-the-art review on deep learning in medical imaging, Front. Biosci. 24 (2019) 392–426.

[15] L. Saba, et al., The present and future of deep learning in radiology, Eur. J. Radiol. 114 (2019) 14–24.

[16] N.A. Alzahab, et al., Hybrid deep learning (hDL)-Based brain-computer interface (BCI) systems: a systematic review, Brain Sci. 11 (1) (2021) 75.

[17] M. Biswas, et al., Symtosis: a liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm, Comput. Methods Progr. Biomed. 155 (2018) 165–177.

[18] M. Biswas, J.S. Suri, Multimodality Imaging of the Heart, Lungs and Peripheral Organs: Deep Learning Applications, IOP Press, 2021.

[19] A. El-Baz, J.S. Suri, Big Data in Multimodal Medical Imaging, CRC Press, 2019.

[20] T. Meng, X. Jing, Z. Yan, W. Pedrycz, A survey on machine learning for data fusion, Inf. Fusion 57 (2020) 115–129.

[21] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, R. Fergus, Regularization of neural networks using dropconnect, in: International Conference on Machine Learning, PMLR, 2013, pp. 1058–1066.

[22] F. Baldassarre, D.G. Morín, L. Rodés-Guirao, Deep Koalarization: Image Colorization Using Cnns and Inception-Resnet-V2, 2017 arXiv preprint arXiv: .03400.

[23] S. Bharati, P. Podder, M.R.H. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images, Informatics in Medicine Unlocked 20 (2020) 100391.

[24] H. Chen, C. Haoyu, Face recognition algorithm based on VGG network model and SVM, J. Phys. Conf. 1229 (1) (2019), 012015 (IOP Publishing).

[25] W. Cheng, Y. Sun, G. Li, G. Jiang, H. Liu, Jointly network: a network based on CNN and RBM for gesture recognition, Neural Comput. Appl. 31 (1) (2019) 309–323.

[26] L. Ding, W. Fang, H. Luo, P.E. Love, B. Zhong, X. Ouyang, A deep hybrid learning model to detect unsafe behavior: integrating convolution neural networks and long short-term memory, Autom. ConStruct. 86 (2018) 118–124.

[27] F. Feng, S. Wang, C. Wang, J. Zhang, Learning deep hierarchical spatial–spectral features for hyperspectral image classification based on residual 3D-2D CNN, Sensors 19 (23) (2019) 5276.

[28] C.A. Ferreira, et al., Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2, in: International Conference Image Analysis and Recognition, Springer, 2018, pp. 763–770.

[29] S. Garg, K. Kaur, N. Kumar, J.J. Rodrigues, Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: a social multimedia perspective, IEEE Trans. Multimed. 21 (3) (2019) 566–578.

[30] Z. Ge, G. Cao, X. Li, P. Fu, Hyperspectral image classification method based on 2D–3D CNN and multibranch feature fusion, IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing 13 (2020) 5776–5788.

[31] A. Gumaei, M.M. Hassan, A. Alelaiwi, H. Alsalman, A hybrid deep learning model for human activity recognition using multimodal body sensing data, IEEE Access 7 (2019) 99152–99160.

[32] M.M. Hassan, A. Gumaei, A. Alsanad, M. Alrubaian, G. Fortino, A hybrid deep learning model for efficient intrusion detection in big data environment, Inf. Sci. 513 (2020) 386–396.

[33] D. He, Z. Yao, Z. Jiang, Y. Chen, J. Deng, W. Xiang, Detection of foreign matter on high-speed train underbody based on deep learning, IEEE Access 7 (2019) 183838–183846.

[34] E.P. Ijjina, C.K. Mohan, Hybrid deep neural network model for human action recognition, Appl. Soft Comput. 46 (2016) 936–952.

[35] N. Jaouedi, N. Boujnah, M.S. Bouhlel, A new hybrid deep learning model for human action recognition, Journal of King Saud University-Computer Information Sciences 32 (4) (2020) 447–453.

[36] Y.-G. Jiang, Z. Wu, J. Tang, Z. Li, X. Xue, S.-F. Chang, Modeling multimodal clues in a hybrid deep learning framework for video classification, IEEE Trans. Multimed. 20 (11) (2018) 3137–3147.

[37] R.M. Kamble, et al., Automated diabetic macular edema (DME) analysis using fine tuning with inception-resnet-v2 on OCT images, in: IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2018, IEEE, 2018, pp. 442–446.

[38] T. Kaur, T.K. Gandhi, Deep convolutional neural networks with transfer learning for automated brain image classification, Mach. Vis. Appl. 31 (3) (2020) 1–16.

[39] Y. Li, S. Chai, Z. Ma, G. Wang, A hybrid deep learning framework for long-term traffic flow prediction, IEEE Access 9 (2021) 11264–11271.

[40] Y. Li, H. Huang, Q. Xie, L. Yao, Q. Chen, Research on a surface defect detection algorithm based on MobileNet-SSD, Appl. Sci. 8 (9) (2018) 1678.

[41] Y. Li, L. Zou, L. Jiang, X. Zhou, Fault diagnosis of rotating machinery based on combination of deep belief network and one-dimensional convolutional neural network, IEEE Access 7 (2019) 165710–165723.

[42] Z. Liu, et al., Hybrid deep learning for plant leaves classification, in: International Conference on Intelligent Computing, Springer, 2015, pp. 115–123.

[43] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, Measurement 167 (2021) 108288.

[44] N. Lu, Y. Wu, L. Feng, J. Song, Deep learning for fall detection: three-dimensional CNN combined with LSTM on video kinematic data, IEEE journal of biomedical health informatics 23 (1) (2018) 314–323.

[45] S. Lu, Z. Lu, Y.-D. Zhang, Pathological brain detection based on AlexNet and transfer learning, Journal of computational science 30 (2019) 41–47.

[46] Y. Ma, Y. Hao, M. Chen, J. Chen, P. Lu, A. Košir, Audio-visual emotion fusion (AVEF): a deep efficient weighted approach, Inf. Fusion 46 (2019) 184–192.

[47] A. Mohan, V.M. Sundaram, V3O2: hybrid deep learning model for hyperspectral image classification using vanilla-3D and octave-2D convolution, Journal of Real-Time Image Processing (2020) 1–15.

[48] A. Mughees, L. Tao, Multiple deep-belief-network-based spectral-spatial classification of hyperspectral images, Tsinghua Sci. Technol. 24 (2) (2018) 183–194.

[49] J.A. Nasir, O.S. Khan, I. Varlamis, Fake news detection: a hybrid CNN-RNN based deep learning approach, International Journal of Information Management Data Insights 1 (1) (2021) 100007.
[50] X.-X. Niu, C.Y. Suen, A novel hybrid CNN–SVM classifier for recognizing handwritten digits, Pattern Recogn. 45 (4) (2012) 1318–1325.
[51] S. Peng, H. Huang, W. Chen, L. Zhang, W. Fang, More trainable inception-ResNet for face recognition, Neurocomputing 411 (2020) 9–19.
[52] H. Polat, H. Danaei Mehr, Classification of pulmonary CT images by using hybrid 3D-deep convolutional neural network architecture, Appl. Sci. 9 (5) (2019) 940.
[53] S.K. Roy, G. Krishna, S.R. Dubey, B.B. Chaudhuri, HybridSN: exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification, Geosci. Rem. Sens. Lett. IEEE 17 (2) (2019) 277–281.
[54] M.U. Salur, I. Aydin, A novel hybrid deep learning model for sentiment classification, IEEE Access 8 (2020) 58080–58093.
[55] A. Shakarami, H. Tarrah, A. Mahdavi-Hormat, A CAD system for diagnosing Alzheimer's disease using 2D slices and an improved AlexNet-SVM method, Optik 212 (2020) 164237.
[56] Y. Sun, X. Wang, X. Tang, Hybrid deep learning for face verification, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 1489–1496.
[57] A.J. Suresh, J. Visumathi, Inception ResNet deep transfer learning model for human action recognition using LSTM, Mater. Today: Proceedings (2020), https://doi.org/10.1016/j.matpr.2020.09.609.
[58] D. Syed, H. Abu-Rub, A. Ghrayeb, S.S. Refaat, Household-level energy forecasting in smart buildings using a novel hybrid deep learning model, IEEE Access 9 (2021) 33498–33511.
[59] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, 2017, 1.
[60] S. Tanberk, Z.H. Kilimci, D.B. Tükel, M. Uysal, S. Akyokuş, A hybrid deep model using deep learning and dense optical flow approaches for human activity recognition, IEEE Access 8 (2020) 19799–19809.
[61] P. Tang, C. Wang, X. Wang, W. Liu, W. Zeng, J. Wang, Object detection in videos by high quality object linking, IEEE Trans. Pattern Anal. Mach. Intell. 42 (5) (2019) 1272–1278.
[62] M. Toğaçar, B. Ergen, Z. Cömert, Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks, Biocybernetics Biomedical Engineering 40 (1) (2020) 23–39.
[63] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G.S. Choi, B.-W. On, Fake news stance detection using deep learning architecture (cnn-lstm), IEEE Access 8 (2020) 156695–156706.
[64] A. Vijayalakshmi, Deep learning approach to detect malaria from microscopic images, Multimed. Tool. Appl. 79 (21) (2020) 15297–15317.
[65] J. wei Tan, S.-W. Chang, S. Abdul-Kareem, H.J. Yap, K.-T. Yong, Deep learning for plant species classification using leaf vein morphometric, IEEE ACM Trans. Comput. Biol. Bioinf 17 (1) (2018) 82–90.
[66] Y. Wu, H. Tan, Short-term Traffic Flow Forecasting with Spatial-Temporal Correlation in a Hybrid Deep Learning Framework, 2016 arXiv preprint arXiv: .01022.
[67] Y. Wu, H. Tan, L. Qin, B. Ran, Z. Jiang, A hybrid deep learning based traffic flow prediction method and its understanding, Transport. Res. C Emerg. Technol. 90 (2018) 166–180.
[68] Z. Wu, X. Wang, Y.-G. Jiang, H. Ye, X. Xue, Modeling spatial-temporal clues in a hybrid deep learning framework for video classification, in: Proceedings of the 23rd ACM International Conference on Multimedia, 2015, pp. 461–470.
[69] J.-L. Xu, S. Hugelier, H. Zhu, A.A. Gowen, Deep learning for classification of time series spectral images using combined multi-temporal and spectral features, Anal. Chim. Acta 1143 (2021) 9–20.
[70] R. Yan, et al., Breast cancer histopathological image classification using a hybrid deep neural network, Methods 173 (2020) 52–60.
[71] X. Yang, Y. Ye, X. Li, R.Y. Lau, X. Zhang, X. Huang, Hyperspectral image classification with deep learning models, IEEE Trans. Geosci. Rem. Sens. 56 (9) (2018) 5408–5423.
[72] S. Yoo, S. Kim, S. Kim, B.B. Kang, AI-HydRa: advanced hybrid approach using random forest and deep learning for malware classification, Inf. Sci. 546 (2021) 420–435.
[73] C. Yu, R. Han, M. Song, C. Liu, C.-I. Chang, A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial–spectral fusion, IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing 13 (2020) 2485–2501.
[74] R. Zhang, Q. Zong, L. Dou, X. Zhao, Y. Tang, Z. Li, Hybrid deep neural network using transfer learning for EEG motor imagery decoding, Biomed. Signal Process Contr. 63 (2021) 102144.
[75] S. Zhang, X. Pan, Y. Cui, X. Zhao, L. Liu, Learning affective video features for facial expression recognition via hybrid deep learning, IEEE Access 7 (2019) 32297–32304.
[76] S. Zhang, S. Zhang, T. Huang, W. Gao, Q. Tian, Learning affective features with a hybrid deep model for audio–visual emotion recognition, IEEE Trans. Circ. Syst. Video Technol. 28 (10) (2017) 3030–3043.
[77] J. Zheng, L. Zheng, A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification, IEEE Access 7 (2019) 106673–106685.
[78] Y. Zhu, X. Gao, W. Zhang, S. Liu, Y. Zhang, A bi-directional LSTM-CNN model with attention for aspect-level text classification, Future Internet 10 (12) (2018) 116.
[79] C. Szegedy, et al., Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[80] E. O. A. https://endnote.com/.
[81] J.S. Suri, S. Laxminarayan, PDE and Level Sets, Springer Science & Business Media, 2002.
[82] C.P. Papageorgiou, T. Poggio, A Trainable Object Detection System: Car Detection in Static Images, 1999.
[83] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012) 1097–1105.
[84] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:.01022.
[85] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[86] G.S. Tandel, A. Balestrieri, T. Jujaray, N.N. Khanna, L. Saba, J.S. Suri, Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm, Comput. Biol. Med. 122 (2020) 103804.
[87] V.K. Shrivastava, N.D. Londhe, R.S. Sonawane, J.S. Suri, Exploring the color feature power for psoriasis risk stratification and classification: a data mining paradigm, Comput. Biol. Med. 65 (2015) 54–68.
[88] U.R. Acharya, L. Saba, F. Molinari, S. Shafique, A. Nicolaides, J.S. Suri, Carotid far wall characterization using LBP, Laws' Texture Energy and wall variability: a novel class of Atheromatic systems, in: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2012, pp. 448–451.
[89] U.R. Acharya, S.V. Sree, P.C.A. Ang, R. Yanti, J.S. Suri, Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals, Int. J. Neural Syst. 22 (2) (2012) 1250002.
[90] U. Acharya, et al., Diagnosis of Hashimoto's thyroiditis in ultrasound using tissue characterization and pixel classification, Proc. IME H J. Eng. Med. 227 (7) (2013) 788–798.
[91] M. Agarwal, et al., Wilson disease tissue classification and characterization using seven artificial intelligence models embedded with 3D optimization paradigm on a weak training brain magnetic resonance imaging datasets: a supercomputer application, Med. Biol. Eng. Comput. 59 (3) (2021) 511–533.
[92] S.-i. Amari, Backpropagation and stochastic gradient descent method, Neurocomputing 5 (4–5) (1993) 185–196.
[93] U.R. Acharya, F. Molinari, S.V. Sree, S. Chattopadhyay, K.-H. Ng, J.S. Suri, Automated diagnosis of epileptic EEG using entropies, Biomed. Signal Process Contr. 7 (4) (2012) 401–408.
[94] U.R. Acharya, S.V. Sree, G. Swapna, R.J. Martis, J.S. Suri, Automated EEG analysis of epilepsy: a review, Knowl. Base Syst. 45 (2013) 147–165.
[95] L. Lin, C. Chen, T. Xu, Spatial-spectral hyperspectral image classification based on information measurement and CNN, EURASIP J. Wirel. Commun. Netw. (1) (2020) 1–16.
[96] J.P. Higgins, J. Savović, M.J. Page, R.G. Elbers, J.A. Sterne, Assessing risk of bias in a randomized trial, Cochrane handbook for systematic reviews of interventions (2019) 205–228.
[97] S. Wang, et al., Performance of deep neural network-based artificial intelligence method in diabetic retinopathy screening: a systematic review and meta-analysis of diagnostic test accuracy, Eur. J. Endocrinol. 183 (1) (2020) 41–49.
[98] M. Biswas, J. Suri, Multimodality Imaging of the Heart, Lungs and Peripheral Organs: Deep Learning Applications, IOP Press, 2021.
[99] U.R. Acharya, et al., Automated classification of patients with coronary artery disease using grayscale features from left ventricle echocardiographic images, Comput. Methods Progr. Biomed. 112 (3) (2013) 624–632.
[100] J.S. Suri, in: Imaging Based Symptomatic Classification and Cardiovascular Stroke Risk Score Estimation, Google Patents, 2011.
[101] U.R. Acharya, K.P. Joseph, N. Kannathal, L.C. Min, J.S. Suri, Heart rate variability, in: Advances in Cardiac Signal Processing, Springer, 2007, pp. 121–165.
[102] U.R. Acharya, et al., An accurate and generalized approach to plaque characterization in 346 carotid ultrasound scans, IEEE transactions on instrumentation measurement 61 (4) (2011) 1045–1053.
[103] J.S. Suri, S. Laxminarayan, Angiography and Plaque Imaging: Advanced Segmentation Techniques, CRC press, 2003.
[104] J.M. Sanches, A.F. Laine, J.S. Suri, Ultrasound Imaging, Springer, 2012.
[105] P. Radeva, J. Suri, Vascular and intravascular imaging trends, analysis, and challenges, IOP Expanding Physics 1 (2020).
[106] M. Maniruzzaman, et al., Accurate diabetes risk stratification using machine learning: role of missing value and outliers, J. Med. Syst. 42 (5) (2018) 1–17.
[107] R. Acharya, Y.E. Ng, J.S. Suri, Image Modeling of the Human Eye, Artech House, 2008.
[108] U.R. Acharya, S.V. Sree, M.M.R. Krishnan, F. Molinari, R. Garberoglio, J.S. Suri, Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan™ systems, Ultrasonics 52 (4) (2012) 508–520.
[109] L. Saba, et al., Automated stratification of liver disease in ultrasound: an online accurate feature classification paradigm, Comput. Methods Progr. Biomed. 130 (2016) 118–134.
[110] U.R. Acharya, et al., Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm, Med. Phys. 39 (7Part1) (2012) 4255–4264.
[111] G. Pareek, et al., Prostate tissue characterization/classification in 144 patient population using wavelet and higher order spectra features from transrectal ultrasound images, in: Technology in Cancer Research Treatment, vol. 12, 2013, pp. 545–557, 6.

[112] R. Narayanan, et al., Adaptation of a 3D prostate cancer atlas for transrectal ultrasound guided target-specific biopsy, Phys. Med. Biol. 53 (20) (2008) N397.
[113] U.R. Acharya, et al., Ovarian tissue characterization in ultrasound: a review, Technol. Canc. Res. Treat. 14 (3) (2015) 251–261.
[114] U.R. Acharya, L. Saba, F. Molinari, S. Guerriero, J.S. Suri, Ovarian tumor characterization and classification: a class of GyneScan™ systems, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012, IEEE, 2012, pp. 4446–4449.
[115] U.R. Acharya, et al., Ovarian tumor characterization using 3D ultrasound, Technol. Canc. Res. Treat. 11 (6) (2012) 543–552.
[116] U.R. Acharya, et al., Evolutionary algorithm-based classifier parameter tuning for automatic ovarian cancer tissue characterization and classification, Ultraschall in der Medizin-European Journal of Ultrasound 35 (3) (2014) 237–245.
[117] A. El-Baz, J.S. Suri, Lung Imaging and Computer Aided Diagnosis, CRC Press, 2011.
[118] G.S. Tandel, et al., A review on a deep learning perspective in brain cancer classification, Cancers 11 (1) (2019) 111.
[119] J.S. Suri, in: Temporal and Spatial Correction for Perfusion Quantification System, Google Patents, 2004.
[120] R.M. Rangayyan, J.S. Suri, Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer, SPIE Publications, 2006.
[121] V.K. Shrivastava, N.D. Londhe, R.S. Sonawane, J.S. Suri, Computer-aided diagnosis of psoriasis skin images with HOS, texture and color features: a first comparative study of its kind, Comput. Methods Progr. Biomed. 126 (2016) 98–109.
[122] A. El-Baz, J.S. Suri, Neurological disorders and imaging Physics, volume 3; application to autism spectrum disorders and alzheimer's, Neurological Disorders Imaging Physics 3 (2019).
[123] L. Saba, J.S. Suri, Neurological disorders and imaging Physics, volume 1; application of multiple sclerosis, Neurological Disorders Imaging Physics 1 (2019).
[124] E. Cuadrado-Godia, et al., Cerebral small vessel disease: a review focusing on pathophysiology, biomarkers, and machine learning strategies, Journal of stroke 20 (3) (2018) 302.
[125] A. El-Baz, J.S. Suri, Machine Learning in Medicine, Chapman & Hall/CRC Healthcare Informatics Series, 2021.
[126] J.S. Suri, in: Mobile Architecture Using Cloud for Data Mining Application, Google Patents, 2014.
[127] F. Bre, J.M. Gimenez, V.D. Fachinotti, Prediction of wind pressure coefficients on building surfaces using artificial neural networks, Energy Build. 158 (2018) 1429–1441.