Deep Learning-Based Model Architecture for Time-Frequency Images Analysis

Haya Alaskar
Computer Science Department
Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia

***Abstract*—Time-frequency analysis is an initial step in the design of invariant representations for any type of time series signals. Time-frequency analysis has been studied and developed widely for decades, but accurate analysis using deep learning neural networks has only been presented in the last few years. In this paper, a comprehensive survey of deep learning neural network architectures for time-frequency analysis is presented and compares the networks with previous approaches to time-frequency analysis based on feature extraction and other machine learning algorithms. The results highlight the improvements achieved by deep learning networks, critically review the application of deep learning for time-frequency analysis and provide a holistic overview of current works in the literature. Finally, this work facilitates discussions regarding research opportunities with deep learning algorithms in future researches.**

Keywords*—Convolutional neural network; time-frequency; spectrogram; scalograms; Hilbert-Huang transform; deep learning; sound signals; biomedical signals

I. Introduction

Time-frequency analysis has been considered for pattern recognition and fault diagnosis. It is usually known as an initial step for signal preprocessing. It provides a suitable tool for analyzing signals in many fields of engineering, biomedicine, finance and speech [1]–[7]. Recently, the importance of discovering powerful signal processing tools has become essential to the analysis of signals. The first time-frequency representation was addressed in the early development of quantum mechanics by H. Weyl, E. Wigner, and J. von Neumann in approximately 1930 [8]. Since then, there have been numerous implementations of time-frequency representation to address signal processing. The early time-frequency analysis system was based on handcrafted techniques. These systems were followed by time-frequency analysis systems based on feature-extraction and machine learning [9]–[11]. Unfortunately, according to a scientist's point of view, preprocessing and feature extraction in any time series signal is not an easy task. There are a number of feature sets that can be extracted from time-frequency domains. Determining the ideal features from such domains requires time for examination and investigation [4], [12]. Furthermore, identify a particular pattern or s of a time-frequency representation is usually unknown [13]. Therefore, effective and reliable tools need to be considered to solve the task. In recent years, studies have been performed to find alternative tools to analyze and identify the pattern directly from a time-frequency image. Starting with the [14] paper, they extracted time-frequency images from the sound signals and used them as input to a deep learning network architecture for classification. Since then, various deep learning network architectures have been proposed, typically based on some form of convolutional neural network (CNN) [10], [11], [13], [15], [16]. In these studies, the CNNs obtain better results than traditional machine learning. Such approaches are attractive since they typically do not need domain knowledge expertise. In fact, CNNs rival human accuracies for the same tasks [17], [18].

Recently, deep learning has proved to be successful in all areas of science, such as successes in image recognition [19], handwriting, manufacturing [13], disease diagnosis [15], [20], [21] and speech processing [22]. The results of these studies have proven the benefits of CNNs in image and signal analysis, which emphasize that CNNs have the capability of addressing diagnosis and classification tasks. Therefore, in the literature, deep learning networks have received considerable attention from researchers; especially, the convolutional neural network. CNNs are able to address data directly without requiring complex preprocessing steps. CNN models are advantageous because of their high levels of expert information processing and can propose much more effective models for complex high dimensional datasets. Therefore, it is important to highlight recent advances techniques of time-frequency analysis, especially recent deep learning architectures, which have outperformed state-of-the-art approaches.

This paper introduces a comprehensive survey of current applications to train a deep learning network in the time-frequency domain in order to classify or diagnose patterns. It will contrast these techniques and compare them among the traditional machine learning applications. To the best of knowledge, this is the first survey that focuses on the use of deep learning with time-frequency analysis and compares it to previous feature-based systems.

The main aim of this survey is two-fold. First, it documents the background knowledge about how the time-frequency domain has been used to address signal processing in the past few years.

Second, it critically reviews the application of deep learning with the time-frequency domain and offers a general overview of the existing literature. In the process of achieving these aims of the paper, the following research questions should be addressed

- Can deep learning be used to classify time-frequency representations of signals?
- Does the deep learning network alter the results of a time-frequency analysis?
- If so, which time-frequency representation of the signal yields the best results?

First, a discussion of the time-frequency representation types and the challenges raised for analyzing the time-frequency domain are presented in section 2. A brief discerption of deep learning networks, especially the CNN, is introduced in section 3. Then, the selection criterion and methodology for selecting which systems to review are explained in section 4. A literature review is highlighted in section 4, and a brief discussion is addressed in section 5.

II. Background

A. Time-Frequency

The time-frequency approach can provide suitable outputs for the discovery of complex, high-dimensional and nonstationary properties. Time-frequency characterization simultaneously represents a signal in both the time and frequency domain. The most popular visual representations of the time-frequency domain are spectrograms and scalograms. This type of representation methods are able to extract particular patterns, for example, the professional extraction of sensitive fault patterns [1]. In medical applications, this type of representation can help to identify an abnormal pattern in biomedical signals. Their success is reported in a number of applications [2]–[7]. Time-frequency methods were also integrated with other advanced algorithms, such as neural networks [5] and support vector machines [8]. In the next sections, a brief introduction is provided about the three types of time-frequency representations.

1) Spectrograms:- a spectrogram is generated using the short-time Fourier transform (STFT). The axis on STFT shows time and frequency, and the color scale of the spectrogram image represents the amplitude of the frequency. The basis for the STFT representation is known as a series of sinusoids.

2) Scalograms:- scalograms are a generated by using the wavelet transform (WT). WTs are a linear time-frequency representation. The basis for the WT representation is a wavelet basis function, which depends on the frequency resolution. The signal is decomposed with different resolutions at different time and frequency scales by scaling and translating the wavelet function.

There are many wavelets types such as the Gaussian, Morlet, Shannon, Meyer, Laplace, Hermit, or the Mexican Hat wavelets. There are differences between each type in both simple and complex functions. There have been many studies to address the effectiveness of each wavelet type. Currently, there is not a clear technique for finding the most suitable wavelet.

3) Hilbert-Huang transform:- the Hilbert-Huang transform (HHT) is considered an adaptive nonparametric representation. It is different from the previous methods such as STFT and WT, which are based on set of basic functions. In contrast, HHT does not need to make assumptions on the basis of the data. It just uses the empirical mode decomposition (EMD) to decompose the signal into a set of elemental signals named intrinsic mode functions (IMFs). The HHT methodology is depicted in Figure 3.

The HHT involves two steps, namely, EMD of the time series signal and the Hilbert spectrum construction. HHTs are particularly useful for localizing the properties of arbitrary signals. For more explanation, see [9].

The HHT does not divide the signal at fixed frequency components, but the frequency of the different components (IMFs) adapts to the signal. Therefore, there is no reduction of the frequency resolution by dividing the data into sections, which gives HHT a higher time-frequency resolution than spectrograms and scalograms.

B. Challenges of Analyzing Time-Frequency Domain

Despite numerous applications using time-frequency representations, analyzing signals have some limitations [10]. Signals usually suffer from several causes of extensive noise, including recording devices, power interference and baseline drift [11]. Hence, the analysis of these signals requires addressing noise and filtering signals.

On the other hand, the features extracted from time-frequency representations need appropriate techniques. Some features can be insufficient to describe the time-frequency domain and will lead to in information loss. In fact, feature selection and extraction expressively need expert knowledge. Furthermore, analyzing time-frequency images to detect features or patterns cannot be accomplished by examining images one by one [1]. Actually, it is very unrealistic to identify a large number of time-frequency images by manual methods. To intelligently and automatically identify the features from many time-frequency images, the prevalent deep learning networks show professional serviceability.

Deep learning is a promising technique for large-scale data analytics[12]. In the literature, they have been used in biomedical signal analysis such as EEG [13], ECG [11], [14]–[16] and EMG [17], [18].

Deep learning networks achieved remarkable result compared with the traditional hand-crafted features. Moreover, once a large size of datasets is available, CNNs are a good method and usually beat human agreement rates. The appearance of deep learning networks has made the analysis of the signals simpler than before.

III. Deep Learning Network (DNN)

DNN is a branch of machine learning tools that has shown significant success in various fields in medicine, business, industry sectors, etc. It attempts to model data hierarchically and classifies patterns using multiple nonlinear processing layers. There are several variants of deep learning such as autoencoders, deep belief networks, deep Boltzmann machines, convolutional neural networks and recurrent neural networks. Since current works have established the success of

CNN deep learning models in the application of time frequency analysis, the concentration of this paper is limited to reviewing the past literatures related to CNN models.

A. Convolutional Neural Network (CNN)

The most successful model of DNN is convolutional neural networks (CNNs). Despite, the CNN was first designated by LeCun et al. in 1998[39]. The golden age of deep learning revolution started when Krizhenvsky et al. [19] won the ImageNet competition by a considerable margin. Since then, only convolutional neural networks have won this ImageNet competition [20], [21].

The differentiation between CNN and the simple multilayer network (MLP) is that MLPs only use input and output layers, and, at most, a single hidden layer, whereas in the DNNs there are a number of layers, including input and output layers [22]. Fig. 1 shows the difference between a simple MLP and a CNN. Each block in the CNN model holds a number of layers.

The CNN contains one or more convolutional and max pooling layers followed by one or more fully connected layers, which perform as the classification layer. Different CNNs employ various algorithms in the convolution layer and subsample layer and different network structures. Finally, the fully connected layers are at the end of the network. In the fully connected layer, weights are no longer shared with the conventional layer. These layers are similar to MLPs, where in the final layer, a SoftMax function is used to generate a distribution over classes.

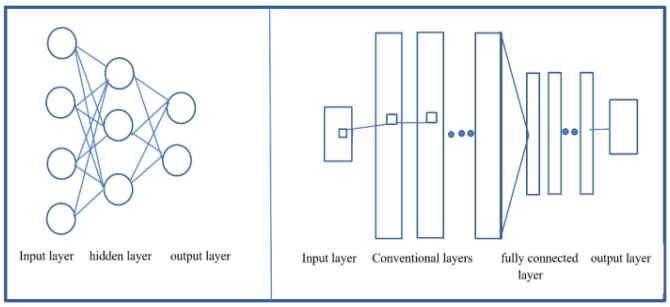

Fig. 1. The Differences in Architecture between a Simple MLP and a CNN.

The significant features of CNNs are that the tasks of preprocessing and feature extraction are not essential in CNNs. In contrast, CNN can automatically identify more complex features because of the number of conventional layers it contains. Furthermore, they are self-learned networks without the need for supervision [35]. This function of DNNs supports the ability of the network to handle large, high-dimensional data that contain a large number of features [36]. This is a beneficial feature of CNNs that reduces the liability during training and helps to select the best features that discriminate classes in the dataset.

IV. Methodology

A. Search Strategy and Selection Process

A database search through online databases such as Google, Google Scholar, and IEEE Explore were used as recommended by [23]. In addition, online databases such as Elsevier, ScienceDirect and ACM, which are the most popular sources for finding scientific papers, were searched. The query terms included time-frequency, DNN based on time-frequency analysis, DNN in signals or time series classification and analysis, etc. also articles that implemented these systems for different languages or domains are included. In total, 154 articles were reviewed and 83 articles were selected for the survey.

B. Literature Sources

The investigation of the applications of DNN with the time-frequency domain was addressed, and articles published in the domain were analyzed.

Most of the selected articles were collected from the publishers, as presented in Table 1, so that the integrity of this review paper is not compromised. However, there is an extensive variety of other sources that are also suitable for this survey.

TABLE I. The Main Sources of the Selected Articles

Publisher
IEEE
Elsevier
bioRxiv
Bioengineering
Springer
Hindawi

C. Data Collection Process

The data collection process involved extensive research of papers that addressed the applications of DNN with time-frequency analysis. These papers were downloaded and studied for collecting suitable information on the subject. The type of results in this paper are qualitative, and the main motivation is to provide a survey of the applications of DNNs and attempt to answer the research questions listed in the introduction section. Overall, the data collection process comprised three main phases

Phase 1: Searching for papers in reliable journals. This phase was completed using some keywords.

Phase 2: Papers are selected and categorized in order to serve the aim of the survey. Then, the qualified papers are examined critically.

Phase 3: Qualitative data were collected and notes were taken to briefly present the data in the results section of this paper. Data were gathered regarding the type of time-frequency domain methods employed.

V. Literature Review

The extensive investigation of the application of DNNs with time-frequency images showed that most of the papers and studies were published after 2016, as represented in Table 2. Most of the papers used the conventional neural network to address this type of image. The next three sections will briefly introduce the applications on DNNs.

TABLE II. THE PUBLISHED PAPERS ON THE APPLICATION OF CNNS ON EEG, ECG AND EMG

References	*Signals and types of representation*	*Deep learning*	*Result*
[24]	EEG, spectrogram	CNN	74% accuracy
[24]	EEG , scalogram	CNN	F1-score 81%
[25]	EEG spectrogram	CNN	96% accuracy
[26]	ECG, spectrogram	DNN	97.5% accuracy
[27]	ECG, Spectrogram	16 CNN	90% sensitivity
[28]	EEG, spectrogram	VGG15	89% accuracy
[29]	facial videos, Spectrogram	VGG15	RMSE was 4.27
[30]	Gait signals, scalogram	CNN	97.06% accuracy
[31]	EMG, spectrograms	CNN	69.23 % accuracy
[32]	ECG, Spectrogram and Hilbert spectrum	CNN	98.3% accuracy
[33]	EEG, spectrogram	CNN	96.16% accuracy
[34]	PPG scalogram	GoogLeNet	92.55% F1
[35]	PCG scalogram	VGG16+SVM	56.2% MAcc
[18]	EMG spectrum	RCNN	90.6 % in R2
[36]	EEG spectrograms	CNN	80% accuracy
[35]	PCG, scalogram image	VGG16	56.2 % accuracy
[37]	EEG, EOG, EMG Spectrogram, scalogram	CNN	95% accuracy
[38]	Sound log-mel spectrogram	CNN	EER was 2.7%
[39]	Sound spectrogram	CNN	71% accuracy
[40]	Sound, spectrograms	VGG	85.36 accuracy
[41]	Sound, spectrograms, scalogram	CNN	74.66 % accuracy
[42]	Sound, spectrogram	CNN	AUC is 0.970
[43]	Fault diagnosis Scalogram	CNN	96% accuracy
[44]	Fault diagnosis, spectrograms	CNN	98%-99%
[45]	Fault diagnosis spectrograms	DNN	95.68% accuracy
[46]	Fault diagnosis Spectrogram	CNN	93.61 % accuracy
[47]	Fault diagnosis Spectrogram, scalogram and Hilbert-Huang.	CNN	81.4%, 99.7% and 75.7% respectively.
[10]	Fault diagnosis, scalogram	PSPP with CNN	99.11% accuracy
[1]	Fault diagnosis Spectrogram	DCNN	96.78% accuracy

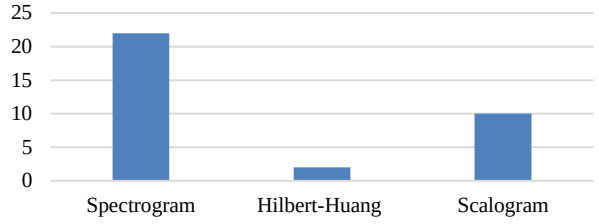

Fig. 2. The Number of Papers used (Spectrogram, Hilbert-Hyang and Scalogram) Type.

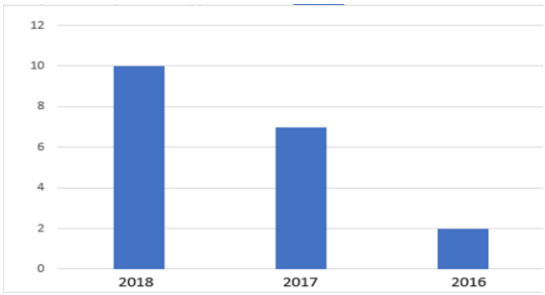

Fig. 3. The Numbers of Papers used Applied the Deep Learning with Time-Frequency Domain on Medical Signals.

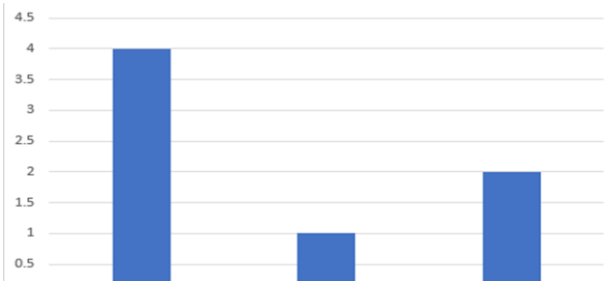

Fig. 4. The Numbers of Papers used Applied the Dnns with Time-Frequency Domain For Fault Diagnosis.

From Fig. 2, it can be noticed that spectrogram has been considered in numbers of studies compared with other type of time-frequency methods. In term of the years of publications. From Fig. 3 and 4, it can be observed that, from 2016 until 2018, considerable effort was undertaken to study and embed the conventional neural network into approaches using these types of data. From the analyzing of different articles , VGG has been selected five times from 31 articles where GoogLeNEt was used only in two papers.

A. Application of CNNs for Fault Diagnosis

Vibration signals are extensively used to diagnose rotating machinery. Researchers attempted to develop automatic and intelligent fault diagnosis tools based on CNN. They extracted the time-frequency representation of vibration signals and fed them directly into a CNN to classify the different kinds of fault features of the rotating machinery. For example, Wang et al. [52] investigate the use of scalogram images as an input to a CNN to predict faults in a set of vibrational data. They used a series of 32 × 32 scalogram images. The highest result they achieved was 96% accuracy.

Lee et al. [53] explored corrupted signals with noise by using a CNN. A short-time Fourier transform was used to generate images from The MFPT data and the Case Western dataset. The trained CNN was able to detect patterns in signals with 98% and 99%.

Janssens et al. [55] incorporated shallow CNNs with the amplitudes of the discrete Fourier transform vector of the raw signal as an input. Pooling, or subsampling, layers were not used. Liu et al.[54] used spectrograms as input vectors into sparse and stacked autoencoders. They attempted to recognize the faults from the normal, inner-race fault, outer-race fault, and rolling bearing parts of fault bearings. The experimental result obtained a good recognition performance on four fault modes with 95.68% accuracy.

Verstraete et al. [18] used a CNN based on time-frequency image analysis for fault diagnosis of rolling element bearings. The CNN consisted of two consecutive convolutional layers without a pooling layer between them. For CNN image inputs, three types of time-frequency transformations are used: short-time Fourier transform spectrogram and wavelet transform (WT) scalogram and Hilbert-Huang transformation (HHT). Their accuracy was 81.4%, 99.7% and 75.7% respectively.

Other study [25] used the Morlet wavelet method to discompose vibration signals of rotating machinery. They used the Pythagorean spatial pyramid pooling (PSPP) layers in the front of the CNN. Hence, the features extracted by the PSPP layer were passed into the convolutional layers for more feature extraction. The evaluation of this model was carried out on two datasets of constant rotating speed signals and variable rotating speed signals. The experiment showed that PSPP CNN was able to achieve 99.11% accuracy.

Another more recent approach in the same manner was proposed in [13]. Xin et al., developed a new CNN to detect different kinds of fault features from the time-frequency representation. The vibration signals were collected from bearings and gears. While the gearbox datasets contain different kinds of faults under the operating conditions, the bearing signals datasets have different fault locations and diameters under several working loads. Those signals are separated into several segments and the time-frequency images are generated by using STFT. These images are treated by the sparse autoencoder method with a linear decoding to expand the sparsity. The proposed DCNN achieved the highest accuracy, with 96.78% compared with the CNN at 89.72% and the LSSVM at 78.33% [13].

B. Application of CNNs for Sound Signals

CNN implementations are becoming more common models in the ASC research domain, where Weiping et al., [50] attempted to use the DCNN for the acoustic scene classification. A CNN model is presented which is similar to the VGG style. They use two types of spectrograms; the first was a generated STFT from raw audio frames, and the second was a CQT spectrogram. The highest result achieved by using the STFT spectrograms images was 0.8536, and the one using the CQT spectrograms images was 0.8052. Weiping et al. conclude that the performance of the CNN could be improved by fine tuning the parameters, normalizing the spectrograms in the training of the DCNN and utilizing the temporal feature.

To better describe sounds that are quite different from speech, Espi et al., [49] used high resolution spectrogram images. These images were directly used as input to a CNN.

However, Thomas et al. [14] used the log-mel spectrogram with its delta and acceleration coefficients to train a CNN. The CNN was evaluated in terms of the SAD accuracy on noisy radio recorded by the Linguistic Data Consortium (LDC) for the DARPA RATS program. Most of the RATS data gained by retransmitting existing audio collections, such as the DARPA EARS Levantine/English Fisher communication telephone speech (CTS) corpus, are broadcast over eight radio channels. In addition, telephone recordings in Arabic Levantine, Pashto and Urdu provided an extensive variety of radio channel broadcast effects.

Other studies conducted to address the efficiency of fusing the mel-scaled short-time Fourier transform spectrogram to train a CNN in [18] determined that using a CNN with the log-mel filter bank energy extracted from the mel-scaled STFT spectrogram outperformed other classifiers. The conclusion of this result was that the log-mel filter bank energy feature possesses fewer coefficients per frame compared to the linear-scaled STFT spectrogram and mel-scaled STFT spectrogram, resulting in a decreased requirement of the parameters of the CNN architecture. In [16], it was asserted that representing audio as images using mel-scaled STFT spectrograms achieved better performance than that achieved with linear-scaled STFT spectrograms, the constant-Q transform (CQT) spectrogram and the continuous wavelet transform scalogram when used as inputs to CNNs for audio classification tasks. The dataset was the ESC-50 dataset, which contains 2000 short (5 second) environmental recordings divided equally into 50 classes. Classes were extracted from five groups, namely, human nonspeech sounds, animals, natural soundscapes and water sounds, exterior/urban noises and interior/domestic sounds. Four frequency-time representations were extracted, namely, linear-scaled STFT spectrogram, Mel-scaled STFT spectrogram, CQT spectrogram, CWT scalogram and MFCC spectrogram. The highest result was obtained by using the mel-scaled STFT spectrogram images, achieving 74.66±3.39 accuracy.

Another novel approach for sound classification of free-flying mosquitoes was proposed by [51]. Their motivation was to detect the existence of a mosquito from its sound signature. A CNN was trained on a wavelet spectrogram. They showed that the CNN performance was better than traditional machine learning classifiers. The result of the ROC analysis was 0.970. The authors concluded that the CNN result was remarkable when compared with traditional feature extraction methods.

C. Application of CNNs for Biomedical Signals

CNN approaches with time-frequency analysis have also been utilized for medical applications. They were employed to serve as decision makers to detect abnormalities in biomedical signals. For example, Hsu et al., [42] used spectrogram images to train a CNN for heart rate estimation based on facial videos.

they have used the GG15 CNN. They claimed that their approach was a novel work that used a DNN network on real-time pulse estimation. They developed a pulse database, named the pulse from face (PFF), and used it to train the CNN.

In [40], spectrogram images were employed to train a CNN for automatic AF detection. The 16-layer CNN was used and achieved 82% accuracy. The CNN recognized normal rhythm, AF and other rhythms with an accuracy of 90%, 82% and 75%, respectively. The conversion of ECG signals to time-frequency images has improved the CNN's ability to automatically perform ECG signal classification, and further, it can also possibly aid robust patient diagnosis.

In this study [39], the time-frequency representations for the heartbeat signal was obtained by using an adapted frequency slice wavelet transform (MFSWT). Features were automatically extracted by the stacked denoising autoencoder (SDA) from the time-frequency image. The DNN classifier was used to identify different pattern on heartbeats. The experiments were applied on the MIT-BIH arrhythmia database. The proposed method gained an accuracy of 97.5%.

Other study [46] investigated if CNNs are able to provide better performance for hypertension risk stratification compared with the traditional signal processing methods. Liang et al., used photoplethysmography (PPG) signals for this investigation. The signals were treated by the continuous wavelet transform via the Morse method to create scalogram images. These images were used to train a pretrained GoogLeNet. The signals included 121 samples from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) Database, and each had arterial blood pressure (ABP) and photoplethysmography (PPG) signals. The classification will be based on blood pressure levels which were normotension (NT), prehypertension (PHT), and hypertension (HT) classes. The experiment was run for the following three trials: NT vs. PHT, NT vs. HT, and (NT + PHT) vs. HT. For the purpose of fitting GoogLeNet, each subject signal was divided into 24 five-second windows. Therefore, 2904 scalogram images were extracted from 2904 signal segments. The F-score obtained to classify NT vs PHT was 80.52%, whereas the approach achieved 92.55% for classifying NT vs HT. The results showed that using a pretrained CNN with scalogram images achieved higher accuracy than that achieved with traditional feature extraction methods.

In [47], the authors examined the ability to train the pretrained VGG16 with scalogram images to classify phonocardiogram (PCG) signals for normal/ abnormal heart sounds. First, the PCG files are segmented into chunks of equal length. Scalogram images are generated using the Morse wavelet transformation. The experimental results showed that the CNN model achieved the highest accuracy at 56.2%, whereas the traditional feature processing with a support vector machine achieved 46.9% accuracy. In total, 3240 PCG signals were collected from 947 pathological patients and healthy subjects.

Gurve and Krishnan [56] employed a CNN on EEG data for classification of the eye state. The spectrogram of the EEG signal is created and fed into a CNN with the NMF features. The implementation of this approach has achieved a good result of 96.16% compared to existing methods for eye state detection.

Eltvik [15] has also applied CNN to analyze the time-frequency domain from EEG signals. He used three types of time-frequency domains. The evaluation of this method involved testing it on two different datasets. The first was an artificial dataset created by simulating a nonstationary and noisy method. The second dataset was real EEG signals made available through the BCI Competition III. It was composed 1,400 EEG signals involving a duration of 3.5 seconds, where each subject was asked to imagine movement in either the right hand or in the left foot. The main task is to identify if the subject was imagining during the experiment. Four different CNN architectures were evaluated using k-fold cross-validation with each of the three representations. The resulting spectrogram and Hilbert spectrum representation of the synthetic data achieved accuracies of 98.3% and 88.19%, respectively. In contrast, the scalogram representation obtained a very poor result of 59.29%. In the real data case, the highest accuracy achieved when classifying the EEG spectrograms was 72.50%. For Hilbert spectra, it was 58.00%, and for scalograms, it was 55.93%.

Ruffini et al. [44] explain how to use a CNN for the REM sleep behavior disorder (RBD) prognosis and diagnosis from an EEG. The EEG data were recorded from 121 idiopathic RBD patients and 91 healthy controls. The signals were taken after a few minutes of being in an eyes-closed resting state. After 2 to 4 years of EEG collecting, 19 of these patients were found to develop Parkinson disease PD and 12 of them had dementia with Lewy bodies, whereas the rest remained idiopathic RBD. Ruffin et al. used a CNN trained with stacked multichannel spectrograms. The performance of a DCNN network reached 80% classification accuracy to classify healthy and PD subjects.

Yuan and Cao [38] attempted to analyze EEGs via spectrogram images by using a CNN. Their motivation was to prove the clinical brain death diagnosis. In this paper Caffe network [57]was used to design a CNN. The EEG signals were acquired from the patients with brain damage. The EEG datasets contained 36 patients, including 19 coma subjects and 17 brain-dead subjects. Spectrogram images were generated from these signals using STFT. In addition, in order to increase the number of created images, six channels of the EEG signals were used to create spectrogram images. In addition, every window of STFT overlapped 20% with the adjacent windows. One hundred spectrogram images were extracted from the EEG data. Based on the experimental result, the CNN was able to distinguish between the coma and brain-dead classes with 96% and 94% accuracy, respectively.

Other researchers shed a light onto how CNNs are able to discriminate sleep stages. For example, [41] used the time-frequency domain of EEG signals in order to classify sleep stages. To reduce the bias and variance in spectrogram images, multitaper spectral estimation was utilized. The dataset included signals collected from 20 young healthy subjects. VGGNet was used with to extracted features by

employ VGG-FE. VGG-FE achieved the highest accuracy with 89%, where most of sleep stages correctly detected slow wave sleep with (89%), rapid eye movement stage (81%), wake stage (78%) and N2 (75%) sensitivity. However, the N1 stage was incorrectly classified with 44% sensitivity.

An analogous study was directed using a CNN for sleep stage detection based on EEGs [37]. In this study, EEGs of 20 healthy young adults were recorded for evaluation. Morlet wavelets were used to produce a time-frequency representation. They achieved a high mean F1-score of 81%, where the accuracy over all sleep stages was 74%.

Andreotti et al. [48] proposed a simple CNN architecture that is trained from scratch using a large publicly available database. They provide EEG, EOG and EMG signals as an input to the CNN. The guided gradient-weighted class activation maps were used for visualizing this network's weights. A large publicly available dataset comprising single night PSG recordings of 200 healthy participants with (STFT). They generated time-frequency transforms for each epoch and modality of the signals. The continuous wavelet transforms (CWT) with a Morlet basis function was used to extract time-frequency images.

Another study was constructed to identify the human gait using the time-frequency representation with a CNN of human gait cycles. For example, [43] used the same approach to detect joint 2-dimensional (2D) spectral and temporal patterns of gait cycles. The signals were acquired from 10 subjects. Each signal was obtained from five inertial sensors that were worn and placed at the lower-back, right hand wrist, the chest, right knee, and right ankle. The experimental results were 91% subject identification accuracy. In this study, they conducted another experiment to improve the gait identification generalization performance by using two methods for an input level and decision score level multisensor combination. The performance improved and the accuracy reached 93.36% and 97.06%, respectively.

Another study attempted to improve CNN performance by combining it with an RNN in order to extract the movement pattern of the upper limb from EMG signals. Xia et al., 2018 [21]. The EMG signals were collected from eight subjects. These signals were recorded in six sessions for each subject and were converted to time-frequency spectrum images and used to train a one-dimensional CNN. The CNN included two recurrent layers in order to develop an RCNN. The experimental result proved that the CNN with the RNN achieved higher accuracy compared with that obtained by using CNNs alone. The authors claimed that these combinations can help to represent the features of EMG signals in the time and frequency domain in a better way. Based on their experimental results, the RCNN model can estimate limb movement with sufficient accuracy, and it was able to extract the features in the frequency domain and was robust against noises.

In this study [48] , the authors proposed the use of the CWT to represent the breathing cycles using scalogram images. The experiment attempted to identify the presence of wheezes and or crackles in breath. The CNN was trained to distinguish the scalograms from different classes. The result showed that the model achieved 84% and 87% accuracy of the class of crackles and wheezes, respectively.

VI. Discussion

The main motivation of this paper was to review various studies and papers that addressed the application of the DNN with the time-frequency representation. After analyzing more than 70 articles, 31 were further examined, and the results of each article were addressed. First, a number of findings were identified, and most of the studies were published during the last three years. In addition, convolutional neural networks, especially CNN that were pretrained, were the most commonly utilized. Furthermore, spectrogram and scalogram images were the most regularly used to train CNNs.

It can be observed that there is a large variety in the type of CNN applications that are used to learn patterns and features from the time-frequency domain automatically. All of the studies have investigated the ability of this approach in medical and manufacturing applications. Each of these studies has confirmed that CNN can extract the optimal information in order to address the required task. Most of these articles' results are comparable to state-of-the-art methods. CNNs are proven to be highly successful in analyzing any signal. Previously, reported studies mainly addressed medical signal analysis and diagnosis with the application of expert-designed features.

For example, a CNN using the time-frequency domain of the presented signals has already been shown to be competitive to traditional approaches. Traditional approaches usually extract a set of features from single or multiple channel signals based on human expertise. Therefore, this could be a difficulty for nondomain experts. Furthermore, traditional feature extraction methods are not capable of utilizing correlation information between different channels. CNNs are very powerful for learning features directly from the time-frequency domain without the need for signal processing and feature extraction methods [49].

Several significant points can be drawn from this survey. Most of the articles obtained their best result without any human intervention. Furthermore, they did not need to have domain knowledge for the analysis of signals. Based on the results of each article, deep learning can be considered as a sound basis for further optimization toward a competitive, fully automated feature extraction method to analyze signals. The potential of directly training a CNN using the time-frequency domain rather than only the time or the frequency domain, for example, in sound signals studies, has been claimed to be related to the time-frequency domain's very detail-rich but sufficiently sparse features that address complex characterization with overlapping sounds [49].

Another important point of this survey was the selection of the STFT-based images to train the CNN, However, studies confirmed that using sclogram is the usually obtained a good result. They motivated by the fact that the scalogram could better represent the nonstationary aspect of any type of signal unlike the STFT. In fact, wavelets are known to provide a robust time-frequency representation for different type of signals as they are localized both in time and frequency.

Therefore, their time–frequency domain information is rich and various [46]. Furthermore, [25] asserted that the wavelet transform is a time-frequency domain analysis tool that offers the best local features of the signal. Because of this, it is frequently used in denoising, feature extraction, and fault diagnosis. Hence, scalogram as input to the CNN can more accurately represent the nature of signals, which improves CNN feature encoding.

VII. Conclusion

This paper is presented to describe the background knowledge of how deep learning has been considered for the field of signal analysis and how it has transformed that field. Then, the state-of-art applications of CNN deep learning models for different types of tasks are identified. Finally, 35 articles from the literature that are related to the field of the study are considered, most of which were recently published since 2016. These articles from the literature are critically studied to provide a general overview on the performance of deep learning models with a time-frequency representation for signal analysis. From the reviews of the outcomes from these studies, it can be concluded that deep learning is able to learn features and patterns directly from time-frequency images. Thus, the brief nature of this survey can make a small but meaningful contribution to the current literature. In addition, it can provide insight on research challenges and future opportunities in the field of signal analysis. Moreover, CNN models generally outperform feature-engineered models.

References

[1] H. Alaskar, A. J. Hussain, F. H. Paul, D. Al-Jumeily, H. Tawfik, and H. Hamdan, “Feature Analysis of Uterine Electrohystography Signal Using Dynamic Self-organised Multilayer Network Inspired by the Immune Algorithm,” in International Conference on Intelligent Computing, 2014, pp. 206–212.

[2] H. Alaskar and A. J. Hussain, “Data Mining to Support the Discrimination of Amyotrophic Lateral Sclerosis Diseases Based on Gait Analysis,” in International Conference on Intelligent Computing, 2018, pp. 760–766.

[3] T. Balli and R. Palaniappan, “Classification of biological signals using linear and nonlinear features.,” Physiol. Meas., vol. 31, no. 7, pp. 903–20, Jul. 2010.

[4] X. Chen, X. Zhu, and D. Zhang, “A discriminant bispectrum feature for surface electromyogram signal classification.,” Med. Eng. Phys., vol. 32, no. 2, pp. 126–35, Mar. 2010.

[5] B. Liu, M. Wang, H. Yu, L. Yu, and Z. Liu, “Study of Feature Classification Methods in BCI Based on Neural Networks.,” in Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2005, vol. 3, pp. 2932–5.

[6] L. O. Machado, “Medical Application of Artificial Network Connectionist Models of Survival,” Stanford University, 1996.

[7] N. K. Orphanidou, A. Hussain, R. Keight, P. Lishoa, J. Hind, and H. Al-Askar, “Predicting Freezing of Gait in Parkinsons Disease Patients Using Machine Learning,” in 2018 IEEE Congress on Evolutionary Computation (CEC), 2018, pp. 1–8.

[8] K. Gröchenig, Foundations of time-frequency analysis. Springer Science & Business Media, 2013.

[9] H. M. Alaskar, “Dynamic self-organised neural network inspired by the immune algorithm for financial time series prediction and medical data classification,” PhD Thesis, Liverpool John Moores University, 2014.

[10] W.-L. Zheng and B.-L. Lu, “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks,” IEEE Trans. Auton. Ment. Dev., vol. 7, no. 3, pp. 162–175, 2015.

[11] H. Zhou et al., “Towards Real-Time Detection of Gait Events on Different Terrains Using Time-Frequency Analysis and Peak Heuristics Algorithm,” Sensors, vol. 16, no. 10, Oct. 2016.

[12] T. Verplancke et al., “A novel time series analysis approach for prediction of dialysis in critically ill patients using echo-state networks,” BMC Med. Inform. Decis. Mak., vol. 10, no. 1, p. 4, 2010.

[13] Y. Xin, S. Li, C. Cheng, and J. Wang, “An intelligent fault diagnosis method of rotating machinery based on deep neural networks and time-frequency analysis,” J. Vibroengineering, vol. 20, no. 6, pp. 2321–2335, 2018.

[14] S. Thomas, S. Ganapathy, G. Saon, and H. Soltau, “Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions,” in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014, pp. 2519–2523.

[15] A. Eltvik, “Deep Learning for the Classification of EEG Time-Frequency Representations,” Master’s Thesis, NTNU, 2018.

[16] M. Huzaifah, “Comparison of Time-Frequency Representations for Environmental Sound Classification using Convolutional Neural Networks,” ArXiv Prepr. ArXiv170607156, 2017.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[18] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, “Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings,” Shock Vib., vol. 2017, 2017.

[19] S. A. Khan and S.-P. Yong, “An Evaluation of Convolutional Neural Nets for Medical Image Anatomy Classification,” in Advances in Machine Learning and Signal Processing, Springer, 2016, pp. 293–303.

[20] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, “Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals,” Inf. Sci., vol. 415, pp. 190–198, 2017.

[21] P. Xia, J. Hu, and Y. Peng, “EMG-Based Estimation of Limb Movement Using Deep Learning With Recurrent Convolutional Neural Networks,” Artif. Organs, vol. 42, no. 5, pp. E67–E77, 2018.

[22] G. Hinton et al., “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82–97, 2012.

[23] S. Shetty and Y. S. Rao, “SVM based machine learning approach to identify Parkinson’s disease using gait analysis,” in Inventive Computation Technologies (ICICT), International Conference on, 2016, vol. 2, pp. 1–5.

[24] N. E. Huang, “Introduction to the Hilbert–Huang transform and its related mathematical problems,” in Hilbert–Huang transform and its applications, World Scientific, 2014, pp. 1–26.

[25] S. Guo, T. Yang, W. Gao, and C. Zhang, “A Novel Fault Diagnosis Method for Rotating Machinery Based on a Convolutional Neural Network,” Sensors, vol. 18, no. 5, 2018.

[26] K. Kim, “Arrhythmia Classification in Multi-Channel ECG Signals Using Deep Neural Networks,” 2018.

[27] M. Längkvist, L. Karlsson, and A. Loutfi, “Sleep stage classification using unsupervised feature learning,” Adv. Artif. Neural Syst., vol. 2012, p. 5, 2012.

[28] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, “Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals,” Comput. Biol. Med., vol. 100, pp. 270–278, 2018.

[29] F. Andreotti, O. Carr, M. A. Pimentel, A. Mahdi, and M. De Vos, “Comparing Feature-Based Classifiers and Convolutional Neural Networks to Detect Arrhythmia from Short Segments of ECG,” Computing, vol. 44, p. 1, 2017.

[30] S. Kiranyaz, T. Ince, and M. Gabbouj, “Real-time patient-specific ECG classification by 1-D convolutional neural networks,” IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675, 2016.

[31] G. Biagetti, P. Crippa, S. Orcioni, and C. Turchetti, "Surface EMG fatigue analysis by means of homomorphic deconvolution," in Mobile Networks for Biometric Data Analysis, Springer, 2016, pp. 173–188.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026–1034.

[33] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning (Book in preparation) -. MIT press, 2016.

[34] C. Szegedy et al., "Going deeper with convolutions," 2015.

[35] S. Savalia and V. Emamian, "Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolution Neural Networks," Bioengineering, vol. 5, no. 2, p. 35, 2018.

[36] U. R. Acharya et al., "Automated characterization of arrhythmias using nonlinear features from tachycardia ECG beats," in 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, pp. 000533–000538.

[37] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel EEG using convolutional neural networks," ArXiv Prepr. ArXiv161001683, 2016.

[38] L. Yuan and J. Cao, "Patients' EEG Data Analysis via Spectrogram Image with a Convolution Neural Network," in International Conference on Intelligent Decision Technologies, 2017, pp. 13–21.

[39] K. Luo, J. Li, Z. Wang, and A. Cuschieri, "Patient-specific deep architectural model for ecg classification," J. Healthc. Eng., vol. 2017, 2017.

[40] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ECG Signal Classification for Detection of Atrial Fibrillation Using a Novel Neural Network," Computing, vol. 44, p. 1, 2017.

[41] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring," ArXiv Prepr. ArXiv171000633, 2017.

[42] G.-S. Hsu, A. Ambikapathi, and M.-S. Chen, "Deep learning with time-frequency representation for pulse estimation from facial videos," in Biometrics (IJCB), 2017 IEEE International Joint Conference on, 2017, pp. 383–389.

[43] O. Dehzangi, M. Taherisadr, and R. ChangalVala, "IMU-Based Gait Recognition Using Convolutional Neural Networks and Multi-Sensor Fusion," Sensors, vol. 17, no. 12, p. 2735, 2017.

[44] G. Ruffini et al., "Deep learning with EEG spectrograms in rapid eye movement behavior disorder," bioRxiv, p. 240267, 2018.

[45] G. Vrbancic and V. Podgorelec, "Automatic Classification of Motor Impairment Neural Disorders from EEG Signals Using Deep Convolutional Neural Networks," Elektron. Ir Elektrotechnika, vol. 24, no. 4, pp. 3–7, 2018.

[46] Y. Liang, Z. Chen, R. Ward, and M. Elgendi, "Photoplethysmography and Deep Learning: Enhancing Hypertension Risk Stratification," Biosensors, vol. 8, no. 4, p. 101, 2018.

[47] Z. Ren, N. Cummins, V. Pandit, J. Han, K. Qian, and B. Schuller, "Learning Image-based Representations for Heart Sound Classification," in Proceedings of the 2018 International Conference on Digital Health, 2018, pp. 143–147.

[48] F. Andreotti, H. Phan, and M. De Vos, "Visualising Convolutional Neural Network Decisions in Automated Sleep Scoring⋆," in ICML Workshop, 2018, pp. 1–12.

[49] M. Espi, M. Fujimoto, K. Kinoshita, and T. Nakatani, "Exploiting spectro-temporal locality in deep learning based acoustic event detection," EURASIP J. Audio Speech Music Process., vol. 2015, no. 1, p. 26, 2015.

[50] Z. Weiping, Y. Jiantao, X. Xiaotao, L. Xiangtao, and P. Shaohu, "Acoustic scene classification using deep convolutional neural network and multiple spectrograms fusion," in Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017), 2017.

[51] I. Kiskin et al., "Mosquito detection with neural networks: the buzz of deep learning," ArXiv Prepr. ArXiv170505180, 2017.

[52] J. Wang, J. Zhuang, L. Duan, and W. Cheng, "A multi-scale convolution neural network for featureless fault diagnosis," in Flexible Automation (ISFA), International Symposium on, 2016, pp. 65–70.

[53] D. Lee, V. Siu, R. Cruz, and C. Yetman, "Convolutional neural net and bearing fault analysis," in Proceedings of the International Conference on Data Mining series (ICDM) Barcelona, 2016, pp. 194–200.

[54] H. Liu, L. Li, and J. Ma, "Rolling bearing fault diagnosis based on STFT-deep learning and sound signals," Shock Vib., vol. 2016, 2016.

[55] O. Janssens et al., "Convolutional neural network based fault detection for rotating machinery," J. Sound Vib., vol. 377, pp. 331–345, 2016.

[56] D. Gurve and S. Krishnan, "Deep Learning of EEG Time-Frequency Representations for Identifying Eye States," Adv. Data Sci. Adapt. Anal.

[57] Y. Jia et al., "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 675–678.