# BRIEF OVERVIEW OF NEURAL NETWORKS FOR MEDICAL APPLICATIONS

Máté HIREŠ, Peter BUGATA, Matej GAZDA, Dávid J. HREŠKO, Róbert KANÁSZ, Lukáš VAVREK, Peter DROTÁR
Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice,
Letná 9, 042 00 Košice, Slovak Republic, Tel. +421 55 602 3175,
E-mail: mate.hires@tuke.sk, peter.bugata@student.tuke.sk, matej.gazda@tuke.sk, david.jozef.hresko@tuke.sk,
robert.kanasz@student.tuke.sk, lukas.vavrek.3@student.tuke.sk, peter.drotar@tuke.sk

**ABSTRACT**

*Neural networks experienced great deal of success in many domains of machine intelligence. In tasks such as object detection, speech recognition or natural language processing is performance of neural networks close to that of human. This allows penetration of neural networks in many domains. The medicine is one of the domains that can successfully harvest methodological advances in neural networks. Medical personnel has to deal with huge amount of data that are used for patients' diagnosis, monitoring and treatment. Application of neural networks in diagnosis and decision support systems have proven to add more objectivity to diagnosis, allow for quicker and more accurate decision and provide more personalized treatment. In this brief review we describe several main architectures of neural networks together with their applications. We provide description of convolutional neural networks, auto-encoders and recurrent neural networks together with their applications such as medical image segmentation, processing of electrocardiogram for arrhythmia detection and many others.*

***Keywords:*** *neural network, convolutional neural network, image segmentation, ECG, U-NET, LSTM, medical imaging*

## 1. INTRODUCTION

The recent methodological improvements in neural networks enabled the penetration of deep neural networks into many application domains. Whether it is recurrent networks, convolutional networks, or other architectures, they quickly found diverse applications and have pushed machine intelligence to further levels. Deep neural networks are able to learn complex models of data representations and perform accurate data analysis.

The computer-based assistance became a crucial part of patient diagnosis, monitoring, and treatment. Whether it is a diagnosis by analyzing medical images [1] or evaluating the acquired medical measurement signals such as electrocardiography [2] or electroencephalograph [3], neural networks are viable sources of supporting information. Furthermore, analyzing electrical health records and unstructured text data can provide information that would be too tedious to gather for human administrators. This represents a significant boost in quality and safety.

For years, one of the domains that have been closely tied with computer vision is medical imaging [4, 5]. The medical imaging community closely follows computer vision development and contributes to the methodology. In addition, medical imaging plays an essential role in patient's diagnosis and monitoring.

Even though there are several success stories of neural networks in the medical domain, several challenges remain to be solved. First, the interpretation and explainability is the inherited disadvantage of neural networks [6]. Some model explainability is very important, especially in medical applications where doctors rely on neural network's decisions. Another obstacle holding back advances is the scarcity of the available data, particularly labeled data. Medical data are very sensitive, so sharing data is usually a complicated procedure that many organizations and administrative bodies closely observe. Additionally, since data labeling needs to be done by highly qualified professionals, it is costly and cannot be easily used by contractors.

This paper outlines several basic neural network architectures and describes their usage in medical applications. In the first part, we establish the theoretical background of different architectures, and in the second part, we focus on various medical domains and the application of the networks in these domains.

## 2. ARTIFICIAL NEURAL NETWORKS

The artificial neural network (ANN) [7] is a computational system inspired by the biological neural networks that make up the brain of animals.

The quintessential model is the feedforward ANN, whose goal in a supervised manner is to approximate a given function $f$. For example, for a classification task, $y = f(x)$ maps an input $x$ to a class $y$. The network defines mapping $y = g(x;\theta)$ and learns the value of the parameters $\theta$, which result in the best approximation of the function $f$.

Feedforward ANNs are typically represented by a composition of many different functions. The model is associated with a directed acyclic graph describing how the functions are composed together. For example, the functions $g^{(1)}$, $g^{(2)}$, $\dots$, $g^{(h-1)}$, $g^{(h)}$ connected in a chain form the function $g(x) = g^{(h)}(g^{(h-1)}\dots(g^{(2)}(g^{(1)}(x)))\dots)$ and constitute the layers of the model, where $g^{(i)}$ represents the $i$-th layer. The overall length of the chain gives the *depth* of the model. The final layer is called the output layer.

During neural network training, the training samples specify directly what the output layer must do at each point $x$; it must produce a value that is close to $y = f(x)$. The behavior of the other layers, called hidden layers, is not directly specified by the training data. The learning algorithm must decide how to use hidden layers to best implement an approximation of $f$. Hidden layers are typically vector valued and their dimensionality determines the *width* of the model. Each element of the vectors is called a node or unit, which resembles a neuron in the sense that it receives input from many other units and computes its own activation value.

The universal approximation theorem states that every continuous function that maps intervals of real numbers to some output interval of real numbers can be approximated arbitrarily closely by a feedforward ANN with just one hidden layer [8, 9]. But the layer may be infeasibly large and may fail to learn and generalize correctly. In many circumstances, using deeper models can reduce the number of units required to represent the desired function and can reduce the risk of overfitting [7]. The advantage of deep neural networks is also their ability to extract high-level, abstract features from raw data. They introduce more complex data representations that are expressed through other, simpler representations. The typical example of a representation learning algorithm is the autoencoder (Section 2.1).

Feedforward ANNs as being simple chains of layers without feedback connections form the basis of many other neural network architectures, often developed for specific tasks [7]. They may be generalized to the recurrent neural networks (RNNs) for sequence processing (Section 2.3). Convolutional networks (CNNs) are a specialized kind of ANN (Section 2.2) for processing data with a known array-like topology, such as time-series data represented by a 1D array taking samples at regular time intervals and image data represented by a 2D array of pixels. In general, the layers need not be connected in a chain. Many architectures build a main chain and then add extra architectural elements to it, such as skip connections from layer $i$ to layer $i+2$ or higher. These skip connections facilitate the flow of the gradient from output layers to layers closer to the input.

Another key factor in architecture design is how to connect a pair of layers to each other. In the default layer described by a linear transformation through a weight matrix **W**, each unit is connected to each unit of the next layer. There are many strategies for decreasing the number of connections, which reduce the number of parameters and the amount of calculations required to evaluate the network, but are often highly problem dependent. For example, CNNs (Section 2.2) use specialized patterns of sparse connections that are very effective for computer vision problems.

### 2.1. Autoencoders

Autoencoder is a type of feedforward neural network used to learn data in an unsupervised manner [10], so it does not need explicit labeled data. The aim of an autoencoder is to learn a lower-dimensional representation (encoding) for a higher-dimensional data, typically for dimensionality reduction [11], just like Principal Component Analysis (PCA), by training the network to capture the most important parts of the input. There is major difference between PCA and autoencoder. PCA uses linear transformations whereas autoencoder uses non-linear transformations. Autoencoder is data-specific. It means that autoencoder is only able to meaningfully compress data similar to what they have been trained on. Since it learns features specific for the given training data, it is different than a standard data compression algorithm. So we cannot expect an autoencoder trained on handwritten digits to compress landscape photos.

Typical autoencoder is composed of two symmetrical sub-models called encoder and decoder. This two sub-models are connected together using latent layer (sometimes called code, bottleneck or latent space). Latent layer is the key attribute of the network design. Bottleneck constrains the amount of information that can traverse the full network, forcing a learned compression of the input data. The figure 1 depicts the simplest architecture of autoencoder with the only one hidden layer.

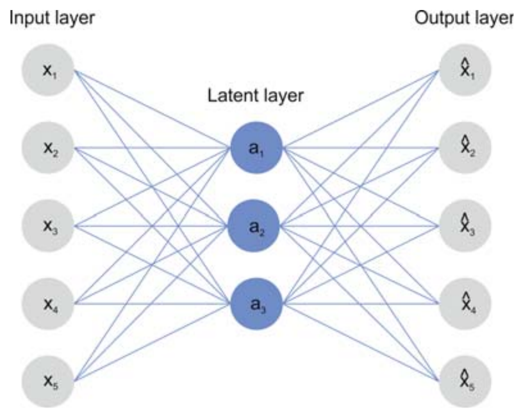


**Fig. 1** Autoencoder with single hidden layer

This simple hidden layer autoencoder has encoder and decoder parts defined as:

$$h = \sigma(W_{xh}x + b_{xh}), \quad (1)$$

$$z = \sigma(W_{hx}x + b_{hx}), \quad (2)$$

where $W$ and $b$ represents weights and biases respectively and sigma is non-linear activation function. The encoder in equation (1) maps input features represented by vector $x$ into a hidden representation $h$ by mapping followed by non-linear transformation using an activation function. On the other hand, decoder defined in equation (2) takes hidden (compressed) representation of input vector and tries to reconstruct it back into the original input space by the same transformation as were used in encoder. The difference between the original input vector $x$ and the reconstructed output $z$ is called reconstruction error. During learning phase the autoencoder tries to minimize reconstruction error. When designing an autoencoder, it is important to pay attention to four different model hyperparameters:

1. Code size: Number of nodes in the middle layer. Smaller size results in more compression.
2. Number of layers: Like all neural networks, an important hyperparameter to tune autoencoders is the depth of the encoder and the decoder. While a higher depth increases model complexity, a lower depth is faster to process.
3. Number of nodes per layer: The number of nodes per layer defines the weights per layer. Typically,

the number of nodes decreases with each subsequent layer in the autoencoder as the input to each of these layers becomes smaller across the layers.

4. Reconstruction loss: The loss function used to train the autoencoder is highly dependent on the type of input and output of the autoencoder.

Autoencoder can be used for a wide variety of applications, but they are typically used for tasks like dimensionality reduction, data denoising, feature extraction, image generation, sequence to sequence prediction, and recommendation systems.

Data denoising is the use of autoencoder to strip grain/noise from images. Similarly, autoencoder can be used to repair other types of image damage, like blurry images or images missing sections.

### 2.2. Convolutional neural networks

Convolutional neural network is a special kind of deep neural network. Here, deep refers to the high number of hidden layers between the input and output layers. CNNs are inspired by the biological visual cortex [12]. Different neurons in the brain responds to different portions of the visual field. This idea of different neurons having different tasks established the CNNs.

CNNs consist of multiple alternating convolution and pooling layers ending with one or more fully connected dense layers. In CNN architectures for classification of RGB images, the input layer has 3 channels, which corresponds to the number of color channels. In hidden layers, number of channels corresponds to the feature maps that encode various attributes of the data. The purpose of these hidden layers is to learn different attributes of the data by extracting certain features from the input data [13]. The convolution layers use convolution operation in their computations in place of multiplication [7]. They use a 3D filter of weights to map the activations from a layer to the next one. The aim of these layers is to compute the feature maps. They are computed by convolving the input with the previously learned kernel and then applying a non-linear activation on the results element-wisely [14]. The feature map values can be computed as

$$z^l_{i,j,k} = \mathbf{w}^{l\,T}_k \mathbf{x}^l_{i,j} + b^l_k, \tag{3}$$

where $z^l_{i,j,k}$ is the feature value at the location $(i, j)$ in the $k$-th map of the $l$-th layer, $\mathbf{w}^l_k$ is the weight vector and $b^l_k$ the bias term of the $k$-th filter of the $l$-th layer, while $\mathbf{x}^l_{i,j}$ is the input on the given position.

The pooling layers task is to reduce the size of the feature maps by decreasing the number of connections between the convolutional layers [15]. The kernels of the first convolutional layers aim is to extract lower level features such as curves or edges, the kernels of higher layers aim to extract higher level features such as shapes or whole objects.

The fully connected layers are responsible for the final classification. In their case, each previous neuron is connected to every neuron of the next layer. Every value is then co-responsible for the final class prediction [16].

They are most suitable for array-like data, such as images or time-series data.

Le Cun et al. first provided a framework of CNN and implemented the LeNet-5 architecture, which can be considered the first CNN network [17, 18]. It could classify handwritten digits and characters. This network architecture could not perform very well, because of the small computing power and lack of training data.

Since the introduction of the LeNet-5 model, many improvements have been made to overcome the shortages of CNNs. Krizhevsky et al. proposed a new deep CNN architecture called AlexNet, which outperformed the present image recognition methods using the ImageNet dataset [19]. From this time, many other architectures were proposed to improve the overall performance, such as the ResNet [20], VGGNet [21] or GoogleNet [22].

Convolutional neural networks showed to be very useful in tasks such as object detection [23, 24], image classification [25, 26], pathological voice analysis [27, 28], handwriting analysis [29, 30], biomedical signal processing [31] or medical image analysis [32, 33] and segmentation [34].

### 2.3. LSTM

Long Short-Term Memory (LSTM) architecture, initially developed by Hochreiter and Schmidhuber in 1997 [35], is a special variant of a Recurrent Neural Network (RNN) architecture. As with standard RNNs, it is designed to process sequences of information of a variable length. It accomplishes it by unfolding itself over time.

Traditional RNNs have a flaw that they can only process sequences with a limited length since they are unable to preserve long-term context due to the vanishing/exploding gradient problem [36, 37]. This limitation is caused by the fact that the same weight matrix is used for all computations unfolded over time, to calculate a hidden state. LSTM address this limitation by preserving a cell state that can be read and modified through sequence processing.

Recurrent neural networks process sequences by iterating through values $x^{(1)}, \ldots, x^{(t)}$, while maintaining a hidden state $s$, based on already seen information. The state $s^{(t)}$ is calculated from the previous sate $s^{(t-1)}$ and current input value $x^{(t)}$ such as

$$s^{(t)} = f(s^{(t-1)}, x^{(t)}; \theta). \tag{4}$$

A single hidden state passed between RNN units is not enough to preserve context and maintain relevant information over an extended time sequence.

LSTM extends on the same principles as RNNs by implementing recurrently connected subnets called memory blocks [38]. Internally, each memory block contains three multiplicative units called gates that control the flow of information passed between memory blocks - input, output, and forget gate.

First, the cell state $s^{(t)}$ is modified by forget gate to remove information that is no longer relevant. Forget gate decides based on previous hidden state and current input

$$s^{(t)} = s^{(t-1)}\sigma(h^{(t-1)} + x^{(t)}). \quad (5)$$

Then, the input gate modifies the cell state $s^{(t)}$ by adding information to it

$$s^{(t)} = s^{(t)} + \sigma(h^{(t-1)} + x^{(t)}) \tanh(h^{(t-1)} + x^{(t)}). \quad (6)$$

Finally, the output gate computes the new hidden state

$$h^{(t)} = \tanh(s^{(t)})\sigma(h^{(t-1)} + x^{(t)}). \quad (7)$$

Using gates, a memory cell can learn which information is relevant and which can be forgotten. This enables it to carry information and maintain context while processing long sequences.

LSTM based networks (and their variations) proven to be useful in several different use-cases like language translation [39], speech recognition [40], and handwriting generation [41]. They are also used for various tasks in a medical domain such as for drug design [42] and depression trend prediction using EEG signal [43].

## 3. MEDICAL APPLICATIONS OF NEURAL NETWORKS

### 3.1. Autoencoders for biomedical segmentation

In general, the image segmentation represents the task, where the main goal is to precisely identify concrete areas of various structures, defined as series of elementary blocks known as pixels (2D domain) or voxels (3D domain). In the context of biomedical segmentation, this task is primarily required by fields, such as radiotherapy, computer-aided diagnosis and investigation of multiple serious diseases. This variety of use cases, results in fact that biomedical segmentation is currently active area of scientific research.

In recent years, the CNNs has been used as primary solution for medical image segmentation and have dominated other traditional methods, such as machine learning methods using hand-crafted features, graph cut methods, shape deformation and variational methods [44]. Another improvement of CNNs was utilization of end-to-end learning handled by fully convolutional networks (FCN), especially by encoder-decoder architecture [45].

The main difference between CNN architecture and FCN architecture is that, FCN is composed of convolutional layers without any fully-connected layer at the end of the network and is capable to spatially transform the feature map of the intermediate layer back to the original size of the input. Thanks to this novel approach, many models with improved architectures have been introduced, such as U-Net named by its U-shaped architecture [45], ResNet with residual blocks [46], DenseNet with densely connected blocks [47], etc.

One of the many applications of mentioned architectures in medical imaging is the segmentation of the liver and its lesion, which are important indicators for various disease diagnosis. Moreover, it is common, that tumors of the abdominal part of the body, can spread to the liver tissue. As the result of these factors, the liver cancer is the second most common cause of cancer death and is the sixth most frequent cancer [48].

Another actively researched application of autoencoders and its variants in medical imaging is the segmentation of morphological structures, such as vessels. Based on the obtained results, authors of [49], proved that usage of autoenconders in the form of the generative adversarial network (GAN) can be successfully applied to segment the retinal blood vessels in fundus images, which are not only used in the screening of retinal vascular diseases but also in the diagnosis of other serious diseases, such as stroke, hypertension and glaucoma.

### 3.2. 1D CNNs for biomedical signal processing

The study of biomedical signals is essential for a wider understanding of physiological phenomena in the human body, whether under normal or pathological conditions. The biomedical signals provide information about the status of body organs and overall health. They are usually captured by devices that measure different types of information, such as brain activity, blood glucose, blood pressure, heart rate, nerve conduction. According to these measurements, physicians decide on diagnosis and treatment.

Cardiovascular diseases are the most common cause of death worldwide. In 2019, they accounted for 32% of all deaths. Cardiac disorders are usually analyzed using the electrocardiogram (ECG) signal, which shows the electrical activity of the heart over time recorded by an electrocardiograph. The cardiac muscle activity is measured through electrical potential differences between electrodes connected to the body surface. A pair of electrodes forms a lead. A conventional twelve-lead ECG uses four electrodes located on the limbs and six electrodes located on the chest surface. However, different monitoring conditions and the need to highlight some anomalies have led to the development of other lead systems that differ in electrode placement or allow the number of electrodes to be reduced.

Although the ECG has long been a method of diagnosing cardiovascular disease, its interpretation requires considerable human knowledge. In recent years, advanced artificial intelligence methods, such as deep learning of convolutional neural networks, have enabled rapid, human-like ECG interpretation. Even signals and patterns that are largely unrecognizable to human interpreters can be accurately detected by deep neural networks [50]. Deploying efficient ECG analysis algorithms on battery-powered portable devices, such as cell phones with wireless ECG sensors, will allow continuous cardiac function evaluation that can be easily integrated into everyday life [51].

Major conditions for the development of automated ECG analysis are public datasets and well-defined procedures that allow the comparison of different algorithms. To ensure these conditions, several large ECG datasets annotated by cardiologists have been published in recent years. Examples are the PTB-XL dataset with diagnostic labels [52] and the MIT-BIH arrhythmia database [53].

Next, we introduce the three most challenging problems in ECG analysis and present some of their solutions using CNNs. Because ECG signals represent time-series data, 1D CNNs are typically used, with the number of channels usually corresponding to the number of leads processed.

### 3.2.1. R-peak detection in ECG recordings

A cardiac cycle is represented in a period of the repetitive ECG signal as the series of waves: P wave, QRS complex, and T wave. R-peak indicates the maximum amplitude in the QRS complex.

The automatic detection of R-peaks in an ECG recording is crucial in the analysis of heart rate variability and the diagnostics of cardiovascular diseases. Many of developed R-peak detection algorithms contain two main parts: the signal processing to enhance the presence of the QRS complex and an amplitude threshold-based peak decision making. However, these algorithms often only perform well with relatively clean ECG signals but are not robust enough to noises and artifacts caused by motions, electrode-skin conductance changes, and the use of dry electrodes [51].

In [54], a deep learning approach for robust R-peak detection in a noisy ECG is presented. The problem is formulated as a regression task that involves obtaining the distance map from the input ECG signal. A distance map (distance transform) is a derived representation of a digital image that labels each pixel in the foreground by the distance to the nearest pixel of the boundary between the foreground and the background. In the method, the R-peaks in an ECG recording are considered analogous to the boundary in images. Here, the distance map provides the distance of each point on the ECG from the nearest R-peak. To predict the distance map, U-Net combined with inception-residual blocks is used. Then, the R-peak locations are easily obtained from the valleys in the distance map.

### 3.2.2. Arrhythmia detection

The term arrhythmia refers to any change from the normal heart rhythm. The heart beat can be too fast, too slow, or irregular. When the heart does not beat properly, it cannot pump blood effectively and the lungs, brain and all other organs may shut down or be damaged. Arrhythmia detection from ECG recordings is usually performed by expert technicians and cardiologists. To automatically detect heart arrhythmias in an ECG, an algorithm must implicitly recognize the different types of waves and distinguish complex relationships between them over time. This is complicated due to the variability in wave morphology between patients as well as the presence of noise [55].

Arrhythmia detection is a sequence-to-sequence task, which takes a sequence of ECG samples $X = [x_1, \ldots, x_k]$ as input, and gives a sequence of labels $r = [r_1, \ldots, r_n]$ as output. Each label $r_i$ can take on one of $m$ different rhythm classes and corresponds to a segment of the input. Together the output labels cover the full input ECG sequence.

The 1D CNN-based algorithm presented in [56] outperforms certified cardiologists in detecting a wide range of arrhythmias from ECG signals recorded with a single-lead wearable monitor Zio Patch [57].

Similar tasks are formulated not only for arrhythmias, but also for other cardiac diagnoses and abnormalities, usually classifying individual heartbeats, which follows the R-peak detection problem. An example of a generally available labeled database for the evaluation of arrhythmia detectors is the often cited MIT-BIH arrhythmia database [53].

### 3.2.3. Automatic ECG annotation

The most challenging problem in ECG analysis is automatic ECG annotation. This is a multilabel classification task, which aims to automatically identify clinical diagnoses from twelve-lead or reduced-lead ECG recordings. The early and correct diagnosis of cardiac abnormalities can increase the chances of successful treatments. Automatic detection and classification of the abnormalities can assist physicians in diagnosing a growing number of ECG recordings.

The research and development of algorithms for automatic ECG annotation have also been supported by popular competitions such as China Physiological Signal Challenge (CPSC) 2018 [58], and PhysioNet/Computing in Cardiology Challenges 2020 [59] and 2021 [60]. In the presented solutions, deep learning approaches were mostly used, including convolutional neural networks in general and ResNet-based approaches in particular [31].

## 3.3. 2D CNN for medical imaging

CNNs can be used in various medical imaging domains, while achieving state-of-the-art performance in classification, detection and segmentation.

### 3.3.1. Classification

Classification in medical imaging is an action or process of classifying an image or part of the image according to shared qualities or characteristics. Neural Network architectures such as VGG-16, residual networks are used, and in case of need fast inference, MobileNet or SqueezeNet [61] can be employed.

Gazda et al. [32] developed a system that classified chest x-rays into healthy subjects and diseased subject. They verified their approach on classification of healthy and patients suffering from Pneumonia, induced by COVID-19 or other diseases.

Esteva et al. [62] developed a 2D neural network for classification of skin cancer, that achieved performance on-par with expert dermatologists.

Diaz et al. [63] utilized neural network for brain tumor classification from MRI images, including meningioma, glioma, and pituitary tumor.

A study performed by explored ability of 2D CNN for classifying tuberculosis on chest radiographs. [64].

### 3.3.2. Detection

Detecting anomalies in medical imaging is a usual task for radiologists. Detection can be also performed for lesion location, lesion tracking or image discrimination. Popular architectures are R-CNN or Fast R-CNN.

Hirasawa et al. [65] used 2D CNN for detection to diagnose gastric cancer in chromoendoscopic images.

Wu et al. [66] utilized an object detection model-ENDOANGEL for real-time gastrointestinal endoscopic examination.

Li et al. [67] incorporated the attention mechanism for detection of glaucoma.

### 3.3.3. Segmentation

Image segmentation is the procedure of dividing a digital image or volume into a multiple disjoint sets of pixels, respectively voxels. For instance, segmentation can be used to select the contours of kidneys on an abdominal CT scan.

Fully connected 2D convolutional neural neural network u-net proposed by [68] revolutionized deep learning segmentatioon. Neural network architecture strongly resembles letter U shape and consists of two parts. The first part is called an encoder, which consists of several alterning convolutional and pooling layers, is responsible of mapping an image to latent space. The second part is called a decoder, which consists of several convolutional and upsample layers, learns a function that finds the contours of the object under observation.

Isensee et al. [69] proposed nnU-Net self-adapting framework for u-net, that heuristically finds important hyperparameters, such as data-augmentation parameters, preprocessing parameters and depth of u-net. nn-UNet has been regularly achieving top-places in medical image segmentation challenges, such as Kidney Tumor Segmentation Challenge [70], Liver Tumor Segmentation Challenge [48], Brain Tumor Segmentation Challenge [71] etc.

## 3.4. LSTM for medical imaging

Long short-term memory recurrent neural networks provide a very useful tool for sequential data processing, such as medical images like MRI pictures or CMR images [72]. The vectorized input is a weakness of the LSTM networks for medical image analysis, because the spatial information is lost. However, the application of convolutional LSTM (CNN-LSTM) proved to be very useful in such cases [73]. CNN-LSTM based models can discover more discriminative spatiotemporal features by replacing the vector multiplication with convolutional operation [74, 75].

LSTMs have achieved state-of-the-art results among others in brain image segmentation [76].

Konwer et al. used an LSTM-based encoder-decoder for the prediction of COVID-19 from chest x-rays (CXR) [77]. They implemented a two-stage LSTM model, which is able to learn both spatial and temporal information from CXR images to model the progression of COVID-19 and predict the lung infiltrate severity. The first module is the LSTM-Spatial module, which aims to learn the spatial dependencies on the image inputs and the LSTM-Temporal module exploits the temporal dependencies between the images from multiple days. For patch extraction a sliding window approach is employed, for image feature extraction, they employed a CNN network.

Zhang et al. have proposed a multi-level CNN-LSTM model for segmentation of left ventricle myocardium in cardiac short-axis cine MR images [78]. The proposed architecture consists of a CNN block and an RNN block. The CNN learns the appropriate features from the image input, which are then downsampled. The RNN block consists of two levels of CNN-LSTM; the output of the RNN block is deconvoluted to obtain the segmentation results. The proposed approach performed better than a CNN model and a one-level CNN-LSTM model.

Gao et al. generalized the LSTM model with a distanced LSTM (D-LSTM) to model the temporal distance with adaptive forget gate and input gate between longitudinal CTs in lung cancer detection [79]. They multiplied a Temporal Emphasis Model (TEM) to the forget gate and the input gate with learnable parameters to incorporate the distance attribute to LSTM. Therefore enabling the TEM to learn across regular and irregular intervals. They validated their approach on three different datasets while achieving improved results compared with baseline LSTM methods. The external validation experiments shows improved generalizability of their proposed model.

Cai et al. improved the pancreas segmentation in abdominal CT and MRI images using CNN-LSTM network [80]. They used a deep CNN followed by a CNN-LSTM and proposed a novel segmentation-direct loss function. The proposed network model is trained to optimize Jaccard index directly therefore generating threshold-free segmentation. They achieved improvements in pancreas segmentation on both CT and MRI images.

Santeramo et al. [81] have proposed two modifications of a time-modulated LSTM for the detection of abnormalities in chest x-rays such as cardiomegaly, consolidation, pleural effusion or hiatus hernia. These LSTM models take into account the time indexes associated to the inputs. All the images of a patient are processed at each time step, First, a CNN extracts the features from the first image. The LSTN takes as inputs the image labels acquired at the previous time-step, the current image features and the time lapse between the two images. For the last image in the sequence, the LSTM predicts the image labels. The other variation of the time-modulated LSTM uses the time lapse only to modulate the internal state. The results denote that the proposed architectures perform better at capturing of the evolution of abnormalities over time. They have proved that the patients x-ray history can be used to improve automated radiological reporting.

Braman et al. [82] analyzed 3D volumes as a sequence of 2D images of annotated volumetric medical imaging data. They utilized an LSTM network to analyze sequentially the image volumes for the presence of emphysema in a portion of scanned region of lung cancer screening low-dose CT images. They first apply convolution for each image slice individually and then a CNN-LSTM is utilized to process the volume slices. The output of the LSTM is a feature set obtained through convolutions with the input and previous slices within the volume. They demonstrate notable model performance in emphysema detection using only weakly annotated image volumes.

## 4. FUTURE RESEARCH AND CHALLENGES

Although recent advancements of deep learning have been astonishing, there still exist many challenges to its application in healthcare.

The main challenge of CNNs application in the medical domain was the scarcity of data. There already exist several methods of reducing the required training dataset size, such as Transfer Learning, Contrastive learning [32] or exhausting data-augmentation techniques. Despite of the recent advances, additional methods can be researched.

Secondly, the real-world application is limited due to its neural network black box behavior, as it does not explain its decision. The explainability is heavily limited, although several works such as GradCAM or deep Taylor decomposition started the trend of research of explainability of deep neural networks.

Many medical objects are represented by non-euclidian data structures, such as drug discovery. Entirely new research direction in neural networks has been started and it's called Graph Neural Networks. Graph Neural Networks are fast gaining popularity [83] and due to it's default rotation and permutation invariance are achieving state of the art performance.

Another research direction based on geometric deep learning [84] is Group Neural Networks. Inventing new rotation and translation equivariant neural networks could be a turning point for the segmentation and classification of anatomical structures in medical imaging.

## ACKNOWLEDGEMENT

This work was supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under contract VEGA 1/0327/20 and by the Slovak Research and Development Agency under contract No. APVV-16-0211.

## REFERENCES

[1] K. S. CHOI and L. SUNWOO. "Artificial Intelligence in Neuroimaging: Clinical Applications". In: *Investigative Magnetic Resonance Imaging* 26.1 (2022), pp. 1–9.

[2] D. S. LARA-MARTINEZ et al. "Artificial intelligence opportunities in cardio-oncology: Overview with spotlight on electrocardiography". In: *American Heart Journal Plus: Cardiology Research and Practice* (2022), p. 100129.

[3] T. B. SHAMS et al. "EEG-based Biometric Authentication Using Machine Learning: A Comprehensive Survey". In: ().

[4] G. CHOY et al. "Current applications and future impact of machine learning in radiology". In: *Radiology* 288.2 (2018), pp. 318–328.

[5] M. L. GIGER. "Machine learning in medical imaging". In: *Journal of the American College of Radiology* 15.3 (2018), pp. 512–520.

[6] A. GHORBANI, A. ABID, and J. ZOU. "Interpretation of neural networks is fragile". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 3681–3688.

[7] I. GOODFELLOW, Y. BENGIO, and A. COURVILLE. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016. ISBN: 9780262035613.

[8] G. CYBENKO. "Approximation by superpositions of a sigmoidal function". In: *Mathematics of Control, Signals, and Systems (MCSS)* 2.4 (Dec. 1989), pp. 303–314. ISSN: 0932-4194.

[9] K. HORNIK, M. STINCHCOMBE, and H. WHITE. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (1989), pp. 359–366. ISSN: 0893-6080.

[10] M. A. KRAMER. "Nonlinear principal component analysis using autoassociative neural networks". In: *AIChE Journal* 7 (Feb. 1991), pp. 233–243.

[11] S. CHAURASIA, S. GOYAL, and M. RAJPUT. "Outlier Detection Using Autoencoder Ensembles: A Robust Unsupervised Approach". In: *2020 International Conference on Contemporary Computing and Applications (IC3A)* (2020), pp. 76–80.

[12] D. H. HUBEL and T. N. WIESEL. "Receptive fields and functional architecture of monkey striate cortex". In: *The Journal of physiology* 195.1 (1968), pp. 215–243.

[13] C. C. AGGARWAL et al. "Neural networks and deep learning". In: *Springer* 10 (2018), pp. 978–3.

[14] J. GU et al. "Recent advances in convolutional neural networks". In: *Pattern Recognition* 77 (2018), pp. 354–377.

[15] A. HYVÄRINEN and U. KÖSTER. "Complex cell pooling and the statistics of natural images". In: *Network: Computation in Neural Systems* 18.2 (2007), pp. 81–100.

[16] M. A. WANI et al. *Advances in deep learning*. Springer, 2020.

[17] Y. LECUN et al. "Handwritten digit recognition with a back-propagation network". In: *Advances in neural information processing systems*. 1990, pp. 396–404.

[18] Y. LECUN et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[19] A. KRIZHEVSKY, I. SUTSKEVER, and G. E. HINTON. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012), pp. 1097–1105.

[20] K. HE et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[21] K. SIMONYAN and A. ZISSERMAN. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).

[22] C. SZEGEDY et al. “Going deeper with convolutions”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 1–9.

[23] W. ZHIQIANG and L. JUN. “A review of object detection based on convolutional neural network”. In: *2017 36th Chinese Control Conference (CCC)*. IEEE. 2017, pp. 11104–11109.

[24] A. DHILLON and G. K. VERMA. “Convolutional neural network: a review of models, methodologies and applications to object detection”. In: *Progress in Artificial Intelligence* 9.2 (2020), pp. 85–112.

[25] T. GUO et al. “Simple convolutional neural network on image classification”. In: *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*. IEEE. 2017, pp. 721–724.

[26] F. SULTANA, A. SUFIAN, and P. DUTTA. “Advancements in Image Classification using Convolutional Neural Network”. In: *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. 2018, pp. 122–129.

[27] L. VAVREK et al. “Deep convolutional neural network for detection of pathological speech”. In: *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. IEEE. 2021, pp. 000245–000250.

[28] M. HIREŠ et al. “Convolutional neural network ensemble for Parkinson’s disease detection from voice recordings”. In: *Computers in biology and medicine* (2021), p. 105021.

[29] G. H. NAIR, V. REKHA, and M. SOUMYA Krishnan. “Handwriting Analysis Using Deep Learning Approach for the Detection of Personality Traits”. In: *Ubiquitous Intelligent Systems*. Springer, 2022, pp. 531–539.

[30] M. GAZDA, M. HIREŠ, and P. DROTÁR. “Multiple-fine-tuned convolutional neural networks for Parkinson’s disease diagnosis from offline handwriting”. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52.1 (2021), pp. 78–89.

[31] L. ANTONI et al. “A Two-Phase Multilabel ECG Classification Using One-Dimensional Convolutional Neural Network and Modified Labels”. In: *Computing in Cardiology 2021* 48 (2021), pp. 1–4.

[32] M. GAZDA et al. “Self-Supervised Deep Convolutional Neural Network for Chest X-Ray Classification”. In: *IEEE Access* 9 (2021), pp. 151972–151982.

[33] Q. LI et al. “Medical image classification with convolutional neural network”. In: *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*. 2014, pp. 844–848.

[34] M. GAZDA et al. “Mixup Augmentation for Kidney and Kidney Tumor Segmentation”. In: *Kidney and Kidney Tumor Segmentation*. Ed. by N. HELLER et al. Cham: Springer International Publishing, 2022, pp. 90–97. ISBN: 978-3-030-98385-7.

[35] S. HOCHREITER and J. SCHMIDHUBER. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

[36] Y. BENGIO, P. SIMARD, and P. FRASCONI. “Learning long-term dependencies with gradient descent is difficult”. In: *IEEE Transactions on Neural Networks* 5.2 (1994), pp. 157–166.

[37] R. PASCANU, T. MIKOLOV, and Y. BENGIO. “On the Difficulty of Training Recurrent Neural Networks”. In: *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*. ICML’13. Atlanta, GA, USA: JMLR.org, 2013, III–1310–III–1318.

[38] A. GRAVES. *Supervised Sequence Labelling with Recurrent Neural Networks*. Vol. 385. Studies in Computational Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. ISBN: 978-3-642-24797-2.

[39] I. SUTSKEVER, O. VINYALS, and Q. V. LE. *Sequence to Sequence Learning with Neural Networks*. 2014.

[40] A. GRAVES, A.-r. MOHAMED, and G. HINTON. “Speech recognition with deep recurrent neural networks”. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013, pp. 6645–6649.

[41] A. GRAVES. *Generating Sequences With Recurrent Neural Networks*. 2014.

[42] A. GUPTA et al. “Generative Recurrent Networks for De Novo Drug Design”. In: *Molecular Informatics* 37.1-2 (2018), p. 1700111.

[43] S. KUMAR and D. SUBHA. “Prediction of Depression from EEG Signal Using Long Short Term Memory(LSTM)”. In: *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. 2019, pp. 1248–1253.

[44] J. PENG and Y. WANG. “Medical image segmentation with limited supervision: A review of deep network models”. In: *IEEE Access* 9 (2021), pp. 36827–36851.

[45] O. RONNEBERGER, P. Fischer, and T. Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab et al. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24574-4.

[46] M. DROZDZAL et al. “The Importance of Skip Connections in Biomedical Image Segmentation”. In: *Deep Learning and Data Labeling for Medical Applications*. Ed. by G. Carneiro et al. Cham: Springer International Publishing, 2016, pp. 179–187. ISBN: 978-3-319-46976-8.

[47] G. HUANG et al. “Densely Connected Convolutional Networks”. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).

[48] P. BILIC et al. “The liver tumor segmentation benchmark (lits)”. In: *arXiv preprint arXiv:1901.04056* (2019).

[49] X. GUO et al. “Retinal vessel segmentation combined with generative adversarial networks and dense U-net”. In: *IEEE Access* 8 (2020), pp. 194551–194560.

[50] K. C. SIONTIS et al. “Artificial intelligence-enhanced electrocardiography in cardiovascular disease management”. In: *Nature Reviews Cardiology* (2021). ISSN: 1759-5010.

[51] M. ELGENDI et al. “Revisiting QRS Detection Methodologies for Portable, Wearable, Battery-Operated, and Wireless ECG Systems”. In: *PLOS ONE* 9.1 (Jan. 2014), pp. 1–18.

[52] P. WAGNER et al. “PTB-XL, a large publicly available electrocardiography dataset”. In: *Scientific Data* 7.154 (2020). ISSN: 2052-4463.

[53] G. B. MOODY and R. G. Mark. “The impact of the MIT-BIH Arrhythmia Database”. In: *IEEE Engineering in Medicine and Biology Magazine* 20.3 (2001), pp. 45–50.

[54] S. VIJAYARANGAN et al. “RPnet: A Deep Learning approach for robust R Peak detection in noisy ECG”. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. 2020, pp. 345–348.

[55] A. Y. HANNUN et al. “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network”. In: *Nature Medicine* 25 (2019), pp. 65–69. ISSN: 1546-170X.

[56] P. RAJPURKAR et al. “Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks”. In: *arXiv preprint arXiv:1707.01836* (2017).

[57] M. TURAKHIA et al. “Diagnostic utility of a novel leadless arrhythmia monitoring device”. English (US). In: *American Journal of Cardiology* 112.4 (Aug. 2013), pp. 520–524. ISSN: 0002-9149.

[58] F. LIU et al. “An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection”. In: *Journal of Medical Imaging and Health Informatics* 8 (2018), pp. 1368–1373.

[59] E. A. PEREZ ALDAY et al. “Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020”. In: *Physiological Measurement* 41.12 (2020).

[60] M. A. REYNA et al. “Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021”. In: *Computing in Cardiology 2021* 48 (2021), pp. 1–4.

[61] F. N. IANDOLA et al. “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 MB model size”. In: *arXiv preprint arXiv:1602.07360* (2016).

[62] A. ESTEVA et al. “Dermatologist-level classification of skin cancer with deep neural networks”. In: *nature* 542.7639 (2017), pp. 115–118.

[63] F. J. D´AZ-PERNAS et al. “A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network”. In: *Healthcare*. Vol. 9. 2. Multidisciplinary Digital Publishing Institute. 2021, p. 153.

[64] P. LAKHANI and B. SUNDARAM. “Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks”. In: *Radiology* 284.2 (2017), pp. 574–582.

[65] Y. IKENOYAMA et al. “Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists”. In: *Digestive Endoscopy* 33.1 (2021), pp. 141–150.

[66] L. WU et al. “A deep neural network improves endoscopic detection of early gastric cancer without blind spots”. In: *Endoscopy* 51.06 (2019), pp. 522–531.

[67] L. LI et al. “Attention based glaucoma detection: a large-scale database and CNN model”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 10571–10580.

[68] O. RONNEBERGER, P. FISCHER, and T. BROX. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[69] F. ISENSEE et al. “nnu-net: Self-adapting framework for u-net-based medical image segmentation”. In: *arXiv preprint arXiv:1809.10486* (2018).

[70] N. HELLER et al. “The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge”. In: *Medical image analysis* 67 (2021), p. 101821.

[71] F. ISENSEE et al. “Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge”. In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 287–297.

[72] A. GRAVES, A.-R. MOHAMED, and G. HINTON. “Speech recognition with deep recurrent neural networks”. In: *2013 IEEE international conference on acoustics, speech and signal processing*. Ieee. 2013, pp. 6645–6649.

[73] Z. YUE et al. "Automatic CIN grades prediction of sequential cervigram image using LSTM with multistate CNN features". In: *IEEE journal of biomedical and health informatics* 24.3 (2019), pp. 844–854.

[74] Y. JIN et al. "SV-RCNet: workflow recognition from surgical videos using recurrent convolutional network". In: *IEEE transactions on medical imaging* 37.5 (2017), pp. 1114–1126.

[75] X. SHI et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting". In: *Advances in neural information processing systems* 28 (2015).

[76] M. F. STOLLENGA et al. "Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation". In: *Advances in neural information processing systems* 28 (2015), pp. 2998–3006.

[77] A. KONWER et al. "Predicting COVID-19 lung infiltrate progression on chest radiographs using spatio-temporal LSTM based encoder-decoder network". In: *Medical Imaging with Deep Learning*. PMLR. 2021, pp. 384–398.

[78] D. ZHANG et al. "A multi-level convolutional LSTM model for the segmentation of left ventricle myocardium in infarcted porcine cine MR images". In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. 2018, pp. 470–473.

[79] R. GAO et al. "Distanced LSTM: time-distanced gates in long short-term memory models for lung cancer detection". In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2019, pp. 310–318.

[80] J. CAI et al. "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function". In: *arXiv preprint arXiv:1707.04912* (2017).

[81] R. SANTERAMO, S. WITHEY, and G. MONTANA. "Longitudinal detection of radiological abnormalities with time-modulated LSTM". In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 326–333.

[82] N. BRAMAN, D. BEYMER, and E. DEHGHAN. "Disease detection in weakly annotated volumetric medical images using a convolutional LSTM network". In: *arXiv preprint arXiv:1812.01087* (2018).

[83] D. JIANG et al. "Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models". In: *Journal of cheminformatics* 13.1 (2021), pp. 1–23.

[84] M. M. BRONSTEIN et al. "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges". In: *arXiv preprint arXiv:2104.13478* (2021).

Received April 20, 2022, accepted June 21, 2022

**BIOGRAPHIES**

**Peter Bugata** received the M.Sc. degree in informatics from the Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice, Slovakia, in 2018, where he is currently pursuing the Ph.D. degree in informatics with the Department of Computers and Informatics. His research interests include feature selection, neural networks and deep learning, and biomedical signal processing. His team won the second place at the 4th China Physiological Signal Challenge 2021, and obtained the fifth place at the 2021 PhysioNet/Computing in Cardiology Challenge, developing the best post-Challenge solution.

**Máté Hireš** received the M.Sc. degree in computer science from the Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice, Slovakia, in 2019. His research interests include neural networks and deep learning, biomedical signal analysis and biomedical decision support systems. He is currently pursuing the Ph.D. degree in computer science.

**Matej Gazda** received the M.Sc. degree in applied computer science from the Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice, Slovakia, in 2019. His research interests include biomedical image analysis, feature selection, and biomedical decision support systems.

**Dávid Jozef Hreško** received the M.Sc. degree in applied computer science from the Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice, Slovakia, in 2021. His research interests include biomedical image segmentation, self-supervised learning, transfer learning and generative adversarial networks.

**Róbert Kanász** received the M.Sc. degree in applied computer science from the Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice, Slovakia. His research interests include machine learning. He is currently pursuing the Ph.D. degree in computer science.

**Lukáš Vavrek** received the M.Sc. degree in applied computer science from the Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice, Slovakia, in 2020. His research interests include neural networks, deep learning and biomedical signal analysis. He is currently pursuing the Ph.D. degree in computer science.

**Peter Drotár** (Member, IEEE) received the M.Sc. and Ph.D. degrees in electronics from the Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice, Slovakia, in 2007 and 2010, respectively. From 2010 to 2012, he was with the Honeywell International, Advanced Technology Europe, Brno, Czech Republic, as a Scientist for Communication and Surveillance Systems. From 2012 to 2015, he was with SIX Research Centre, Brno University of Technology, Brno, as a Postdoctoral

Research Assistant. He is currently an Associate Professor with the Department of Computers and Informatics, Technical University of Kosice. He leads research and development projects concerning biomedical decision support systems. His research interests include biomedical signal and image processing, feature selection, and pattern recognition.