Fluid Temperature Detection Based on its Sound with a Deep Learning Approach

Arshia Foroozan Yazdani[1]
[1]Allameh Helli Tehran College, Tehran, Iran
Email: [1]arshiaforoozanyazdani@yahoo.com

Ali Bozorgi Mehr[2], Iman Showkatyan[3], Amin Hashemi[4], Mohsen Kakavand[5]
[2,3,4]Allameh Helli Tehran College, Tehran, Iran,[5] Department of Computing, Sunway University, Bandar Sunway, 47500, Malaysia
Email: {[2]ali.bozorgi.mehr, [3]Iman.shokatian, [4]Aminhashemi92}@ gmail.com, [5]mohsenk@sunway.edu.my

Abstract: The present study, the main idea of which was based on one of the questions of I.P.T.2018 competition, aimed to develop a high-precision relationship between the fluid temperature and the sound produced when colliding with different surfaces, by creating a data collection tool. In fact, this paper was provided based on a traditional phenomenological project using the well-known deep neural networks, in order to achieve an acceptable accuracy in this project. In order to improve the quality of the paper, the data were analyzed in two ways:

I. Using the images of data spectrogram and the known V.G.G.16 network.
II. Applying the data audio signal and a convolutional neural network (C.N.N.).
Finally, both methods have obtained an acceptable precision above 85%.

Index Terms: Fluid temperature; data spectrogram; V.G.G.16 (Visual Geometry Group); C.N.N (Convolutional Neural Network); the environmental sounds; sound classification; physical-computational research; Deep learning

1. Introduction

Understanding the environmental sounds, including fluid sounds, is regarded as a forward step towards building environmental smart system because the environmental sounds form a wide range of our everyday voice events. Therefore, understanding the fluid sound, as a part of the surrounding voice events, plays an important role in solving the problem of environmental sound classification. In addition, the observations of this project have shown that an approximation of the temperature of a fluid is one of the information derived from fluids sound. Thus, analyzing a fluid sound makes it possible to obtain an approximation of fluid temperature, in addition to understanding a part of the environmental sounds.

Nowadays, deep learning, as a branch of machine learning, has provided significant advantages in various fields, including sound analysis. Further, the use of deep learning has considerably increased compared to the previous periods, considering the growing power of computer hardware [1,2]. This information networking method has been transformed from mathematical theory to a practical macro-data analysis system and used at national managerial levels to a simple Google search. This method has commonly used in different fields and today, even the public are using technologies derived from this method, either knowingly or unknowingly.

However, using this method at the community level does not necessarily mean that scientific and academic institutions do not use this method in their research because it has been highly considered in analyzing laboratory data from the beginning. For example, the high-level projects of recent years have utilized this method or its corresponding methods in their analysis section. Analyzing a massive volume of data over a little period of time, eliminating human error during the data analysis process, and transforming the data abstract properties into analytic features are among the reasons, which have often led researchers to use this method. In this regard, the physical-computational research carried out with this approach are as follows: quantum key distribution [3,4], quantum error correction [5], wave function reconstruction and optimization [6], improving the efficiency of quantum annealing [7], sampling from many-body systems [8], analyzing phase-transition phenomenon [9-11], quantum tomography [12], designing quantum experiments [13,14], wave propagation analysis [15], quantum topology [16], surface codes analysis [17], analyzing Fermi-Hubbard model [18], and analyzing superconducting qubit [19].

Regarding the above-mentioned properties and achievements, this method was used in this project to enhance its scientific grade and technical accuracy.

Deep learning is a subset of the feature learning while feature learning is a subset of machine learning, and ultimately, machine learning is a subset of deep learning (Fig.1).
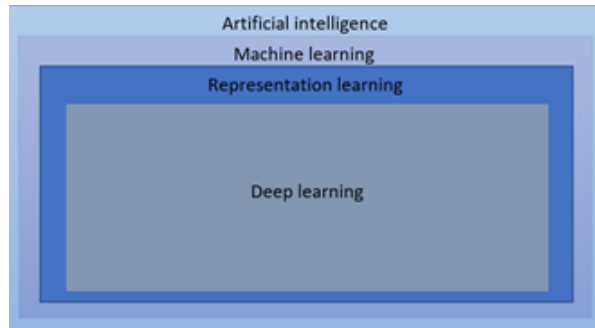

Fig.1. Venn diagram of artificial intelligence

In general, machine learning means linear or nonlinear modeling by artificial neural networks. More specifically, the mechanism of machine learning is such that a set of tagged (classified) data (words, images, and sounds) with certain features are sent to the network layer by layer, and then, the network is trained using these data (if new data is given to the network, it determines the tag type of that data depending on its features). In each layer, the network estimates the tag type of data based on the given features, and attributes one percentage of the error to itself according to the obtained results (the loss function determines this percentage error) and corrects itself to optimize the results (reaching the lowest possible error percentage). In other words, the network behaves data as an optimization problem.

In shallow learning, which is the opposite of deep learning, it is necessary to first specify the features of data sent to the network. Before using the network, the programmer should search for patterns in data to use them in the network. Given the specified features, the network only attempts to create a relationship with tags, resulting in intensifying the workload of the programmer and increasing the probability of human error. However, in the deep learning, the network layers search all data to find features related to the tags and then, transfer them to the classifier, in order to present estimations about data tags through their combination. Thus, it is no longer necessary to previously determine the features in data by the programmer [20].

In deep learning, the workload of the programmer is less, and the network does more work and consequently, has more complexity (Fig.2).
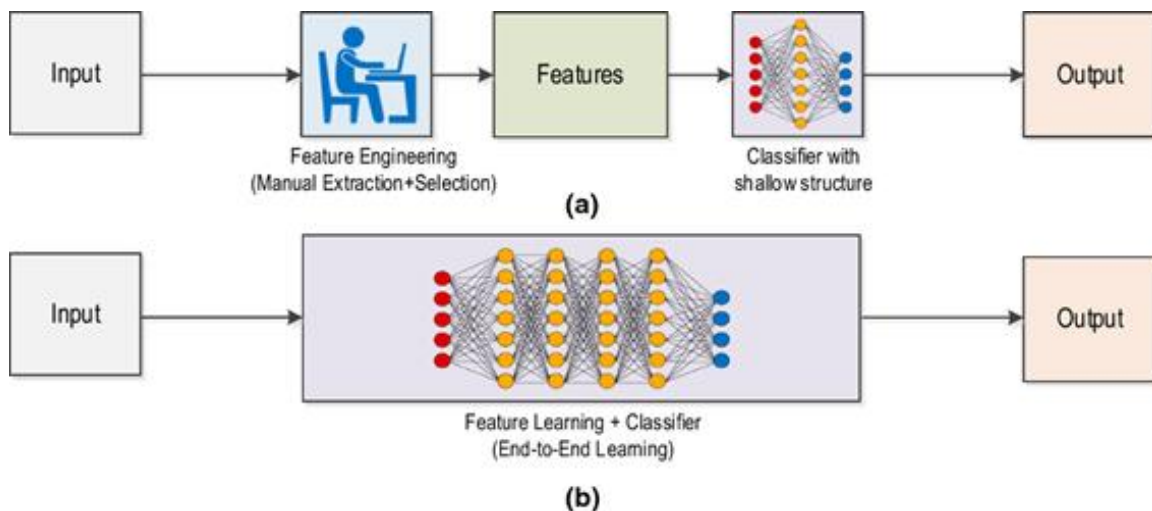

Fig.2. The difference between (a) shallow learning and (b) deep learning [21].

This action in deep learning is performed by increasing the number of data and layers (more layers mean more depth). In fact, increasing the number of data and layers provides the ground for obtaining a more accurate network (in the other types of machine learning, an increase in data leads the accuracy to a certain amount, which does not exceed a specified limit. (Fig.3) shows the fact that the performance of traditional machine learning methods plateaus while the performance of deep learning does not as we add more data [21]. For this reason, processing deep learning requires powerful hardware and more data compared to other subsets of machine learning.
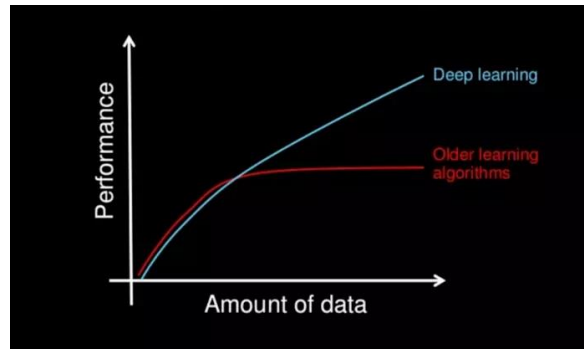
Fig.3. The difference between the performance of deep learning and the other types of learning algorithms with increasing data [22].

In summary, the use of this method slightly increases the final cost of the project, which is why other subsets of the machine learning such as shallow learning are used during some scientific projects, although deep learning has been preferred in this project to increase the quality of work.

It should be noted that the accuracy resulted from increasing data in deep learning is at a higher level with respect to the other types of machine learning.

Technically, the analysis of sounds (or the analysis of any sequential linear or nonlinear time dependency such as films) in machine learning is performed using time-dependent networks (networks dependent on the order of data) in such a way that the features of each data depend on the features of the previous data. This process can be implemented by creating short-term memory, although the volume of this memory is different in various networks.

The "long short-term memory architecture", as one of the most popular architectures of time-dependent networks, along with convolutional architecture are two of the most important and popular neural networks. The importance of "long short-term memory" architecture is that most networks had the capacity of up to 10 time-step before the emergence of networks with this architecture in 1997. However, the presentation of this architecture created a huge difference in the capability of communication between data with hundreds of steps, and hence, led to the significant use of time-dependent networks [20]. Recurrent neural networks are another type of time-dependent networks, which allow converting a sequence with unknown length to a vector with constant magnitude, while maintaining many of the syntactic and structural properties of the input sequence. Some researchers believe that these networks are the most important innovation of neural networks in the field of natural language processing.

In general, the recurrent neural network is a function that receives an input with non-constant magnitude (such as a sentence) as a sequence of n d_{in}-dimensional vectors and returns an output vector y with d_{out} dimension:

$$y_n = RNN(x_{1:n})$$
$$x_n \in \mathbb{R}^{d_{in}}, y_n \in \mathbb{R}^{d_{out}} \tag{1}$$

where $x_{1:n}$ represents the input sentence, whose words are mapped and stacked together with embedding vectors. Additionally, the output y_n can be used at various tasks; for example, it is possible to obtain a representation y_n for each data and give it to the classifier as an input, in order to train the network [23-26]. Nevertheless, the networks used in this project are convolutional and VGG16 networks.

The convolutional neural networks consist of neurons with learnable (adjustable) weights and biases. Each neuron receives several inputs and then, calculates the product of the weights in the inputs, and finally, presents a result using a non-linear transformation (activation) function. The whole network proposes a differentiable score function so that the raw pixels of the input image are located on the one side while the scores related to each class are on the other side. These types of networks still have a cost function such as SVM and Softmax in their mostly fully-connected layer [27].

VGG16 was introduced by the researchers of the "Visual Graphics Group" in Oxford. This network is known for its pyramidal shape, in which the layers close to the image are wider while the farther layers are deeper. VGG16 network consists of a series of convolutional-computational layers, behind which are pooling layers that shrink the layers [28].

Despite the recent development of many time-dependent networks, deep learning is mainly focused on image processing and some of the studies conducted in this area are distinguishing and categorize driving signs [29], house numbers [30], handwritten digits [31], CT scan [32], and electron microscopic images [33]. Considering the sound processing field, few studies have been conducted in this field, especially classifying and grading fluids sounds, except the analysis and recognition of human voices, which is a highly active area and works such as speech [34-40] recognition have been performed in this regard. However, studies have been recently conducted on the classification of fluids sounds in other areas, including human heartbeat [41], urban sounds [42-48] (natural sounds such as children's

play sounds, car horns, air conditioning sound, engine sounds, etc.), and music [49-56], using deep learning. To the best of our knowledge, no article has been published in the field of classifying and identifying sounds created by fluids, by applying deep learning. As a result, the present study can be considered as an innovative article in this regard.

Given the relationship between fluid temperature and the sound created by the fluid, it can be said that the quantity of dynamic viscosity (μ/P) is the intermediate between these two parameters. This quantity, which examines the adhesion and friction between the particles of a material in a macroscopic way, is related to the fluid temperature based on Equation 2. addition, dynamic viscosity has a relationship with the sound created by the fluid, due to its impact on the number and velocity of the waves and bubbles after the fluid collision with the surface [57].

$$\mu(T) = 2.414 \times 10^{-5} \times 10^{247.8/(T-140)} \tag{2}$$

The following cases are among the applications derived from this relationship:

Determining the fluid temperature (Industrial example: Determining the temperature of the molten metals in the iron melting furnaces using this relationship, instead of applying a laser thermometer, due to its lower price and higher accessibility (mobile phones)).

Specifying the dynamic viscosity of the fluid by Equation 2 (this method is less expensive, more accessible, and faster than the conventional laboratory methods such as Brookfield viscometer.

Defining the purity of fluids (an increase in fluid impurities changes the dynamic viscosity of the fluid depending on the type of impurities)

Characterizing the temperature of fluids in natural environments and with few facilities such as waterfalls and rivers

Determining the temperature of fluids, in which the conventional thermometric methods cannot be used due to the high temperature or acid-alkaline nature (Specific example: The laser thermometer fails to determine the temperature of the transparent bodies).

2. Materials and Methods

A. Data Acquisition

In this experiment, the data acquisition process was performed using the device shown in (Fig.4). This device consisted of:

- An aquarium of glass type with a base area of $6.27dm^2$ and a height of $22.5cm$ (total volume of $14.1075lit$), where the water outlet is located at its bottom face at a distance $119.5cm$ from the inclined surface (water drain);
- An inclined surface with an angle of $29.7448°$ is of aluminum type with an area of $284cm^2$, which is $23.0cm$ below the water outlet from the pipe and inside the acoustic environment;
- An acoustic environment, in order to reduce the voice interference at the water drain location
- Heating systems and thermometer with an accuracy of $\pm0.1k$ and model "HANYOUNG ED6-FPMAP4", which can operate at temperatures between -100 to $400°C$ and have an IP68 standard;
- The water transfer system (pipe) with a length of $2m$, a diameter of $7.2\pm0.05mm$ and the skin thickness of $1\pm0.05mm$, which is drawn from the aquarium (heating place) to the acoustic medium special for sound recording;
- Microphone with model "Huawei stereo 0178" at a distance $38cm$ from the point of water collision or the inclined surface.

The mechanism of device operation is such that $3.5lit$ of water in the aquarium reached the desired temperature using the available heating system and then, transferred to the sound recording acoustic environment through water transfer system, in order to be left on the inclined surface. The sound of water collision with the inclined surface was recorded by the microphone and tagged as the main data of the project.

Using this device, 765s data were recorded with the format m4a and in two channels in an asymmetric way (inequality of the number of members in each group) at 6 temperature groups. The interval between each of the two groups was a temperature with a distance of $10°C$ and all of these groups were in the temperature range of $20.8°C$ to $72°C$.

The limitation of the temperature range is due to the following reasons:

The temperature cannot be less than $20.8°C$ because the water spillway is at a lower height to minimize the error of the sound of droplets created by water collision with the inclined surface. For this reason, cold water cannot be converted into drops due to the high adhesion of its droplets.

The temperature cannot be higher than $72^{\circ}C$ because of the limitation of the heating system and heat loss in the aquarium, which prevent the temperature from rising.
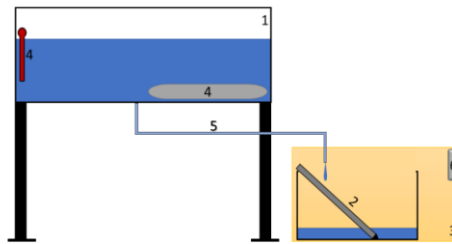

Fig.4. The configuration of the test setup; a pictorial guide to the explanations

B. General Pre-processing

Implementing deep learning algorithms require grouping and tagging data by default, as well as determining the type of using any data by the network. The network used in this project is not an exception from this rule, and accordingly, 80% of the whole 765 data were assigned to the training data and 20% of the remaining data was assigned to the test data.

Meanwhile, in this project, the data of each category were recorded sequentially and continuously to have the least difference with each other (the main difference of the data should be only the type of their category). Then, the whole data were divided into separate data with a length of 1s for utilization in the network. The audio format was also converted from m4a to wav in order to reduce the volume of network processing (due to the change from encrypted file to unencrypted one).

Further, since the type and mode of noise distribution are not the same in different data and the network may consider the noises in any data as an attribute for that data, a normal noise (with a normal distribution in each data) was added to all data. This noise is visible in all data and therefore, the network fails to learn it because it cannot use the noise to differentiate between data.

C. Network Training

This research has addressed the classification of fluids with different temperatures based on their sound, by using the two following approaches:

1) Classification based on the spectrogram image [47,58]
2) Classification based on the raw audio signal

1) Classification based on the spectrogram image:

In order to evaluate an audio signal from the viewpoint of the spectrogram, each sound can be considered as an image, and the problem of sound classification should be converted to the problem of image classification. To this aim, different well-known networks such as VGG16 (which is used in this article), VGG19, and ResNet can be used to follow the process.

The spectrogram of a time signal is obtained by a short-time Fourier transform (STFT). After acquiring the signal spectrogram, the dimensions of the spectrograph images will change to 224×224 to make it suitable for entering into the VGG16 network.

The standard form of displaying the audio signal is such that its horizontal axis is the time in seconds and its vertical axis is in Hertz (Hz).

This chart is data with a temperature of 20.8c $^{\circ}$ for the 10th second (Fig.5).
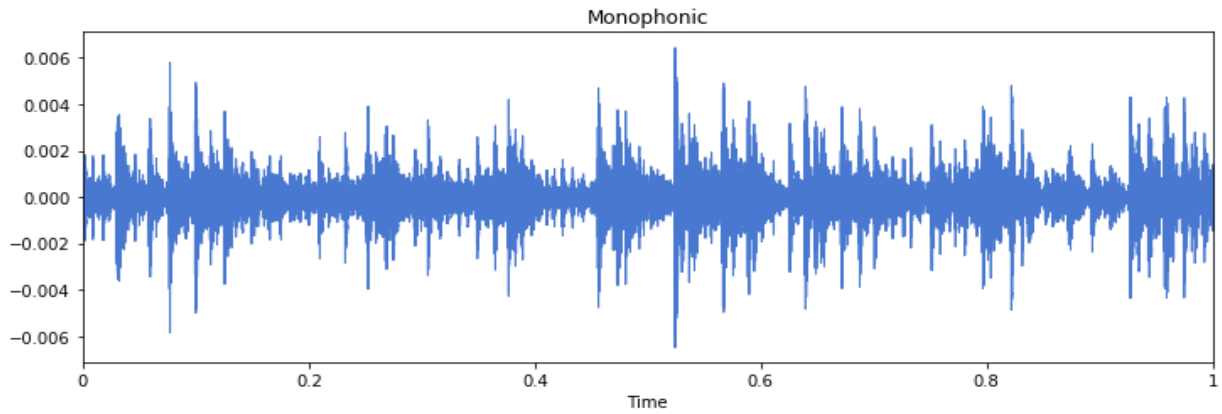
Fig.5. Raw audio signal (Magnitude-Time)

The blue signal is related to the Percussive chart rate while the red signal is for the Harmonic chart rate
The 10th second is data with a temperature of 20.8c °(Fig.6).
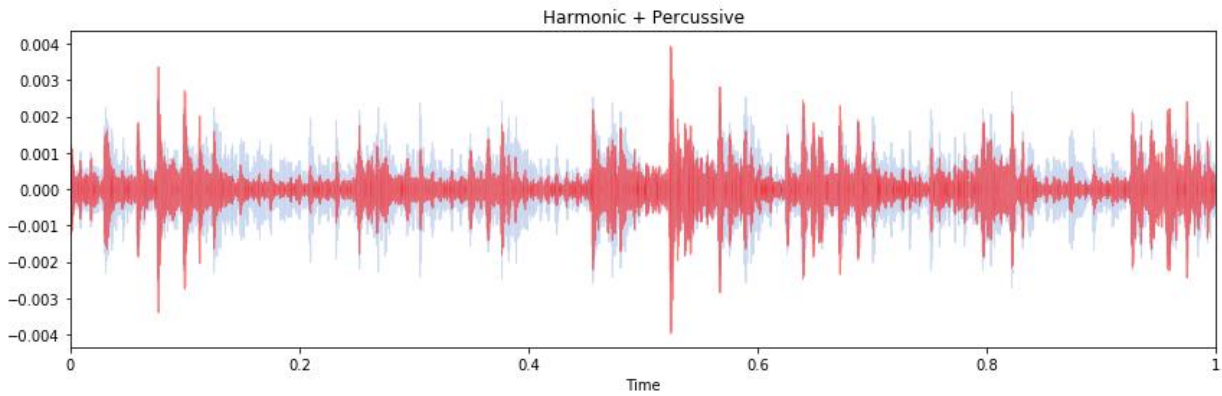
Fig.6. An audio signal with two different chart rates

The low and high spectrograms are for the Percussive and Harmonic chart rates, respectively. Spectrograms are obtained by Python programming (Fig.7).

Fig.7. Two spectrograms of data with two different chart rates

This chart is data with a temperature of 20.8 $^\circ C$ for the tenths second.

Before entering data into the network, a pre-processing step is performed on the images of the spectrograph as follows:

- Image enhancement using contrast-limited adaptive histogram equalization (CLAHE)
- Data Augmentation such as: Crop, mirror, mini-batch strategy

The model was trained using the Adam algorithm, Batch size 20, with 4500 Epochs. The initial learning rate was set to 0.001 and in each 500 Epoch, the value of 0.5 was reduced from the learning rate. In this study, the categorical cross-entropy cost function was used due to the 6 classes of the problem (Fig.8).

Fig.8. (59) VGG16 Network Architecture

Second place in the I.L.S.V.R.C 2014 competition, 13 convolutional layers, 3 fully-connected layers, and 5 "max pooling" layers with dimensions 2×2 [28,59].

This layer divides the image (matrix) into pieces with 2×2 dimensions and inserts the maximum value (determined based on the weights) between four parameters of each piece, instead of the entire piece (to reduce the computations) [61] (Fig.9).

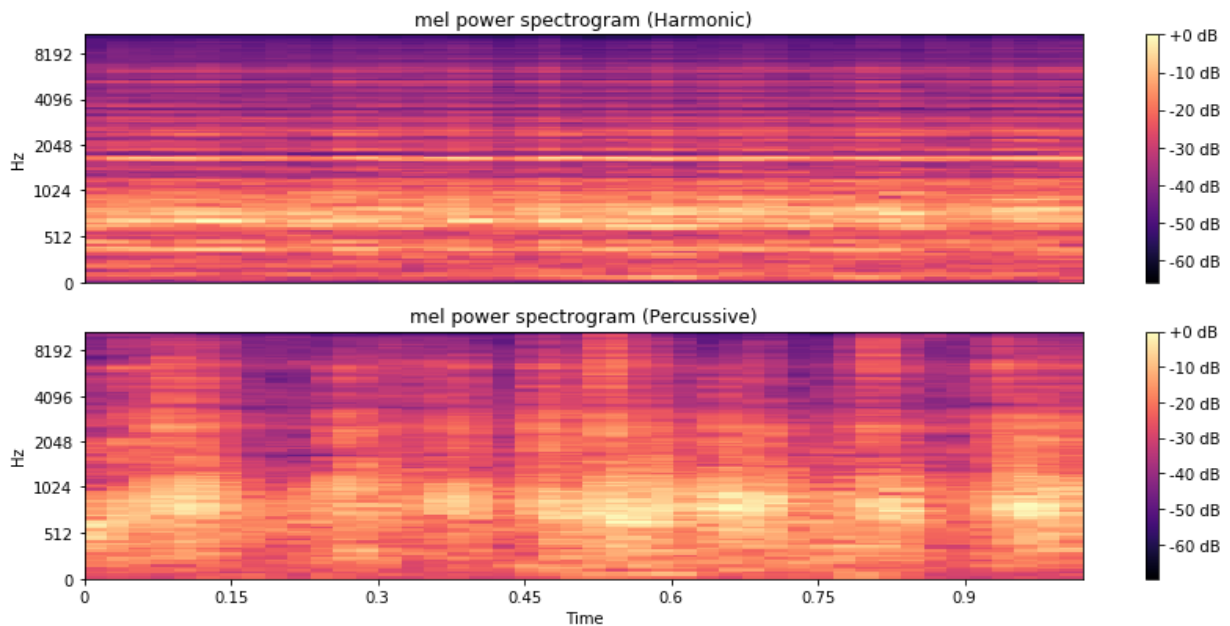Fig.9. [60] Illustrating the function of "max pooling" layer with dimensions 2×2

2) Classification based on the raw audio signal:

Considering the classification of data based on the raw form of the audio signal, the data were entered into a one-dimensional convolutional network (below architecture) as a tensor with dimension (48000×1×1). The dimensions of width and depth were added to the data in terms of dummy axes, in order to make possible the use of the convolutional operator in networks. Before entering data into the network, a series of preprocessing operations were performed on the data as follows:

- Normalizing data
- Standardizing the mean and standard deviation of all data

Then, the data were entered into the convolutional network, which is called the Thermo Fluid Convolution Neural Network (TFCNN), based on its convolutional architecture and its function, i.e. detecting the temperature of a fluid. The "categorical Cross-entropy" cost function was used to train this network. Further, we used the stochastic gradient descent optimization algorithm with the training rate 0.001 and the number of Epoch 30000, which was reduced from the learning rate 0.1 in each 2000 Epoch.

4 convolutional layers, 1 global average pooling layer, and 1 max pooling layer with dimensions 2×2.

The features of all tags (classes) are combined together in fully-connected layers. However, the features of each separate tag are given to the cost function in the "global average pooling" layer [62] (Fig.10).
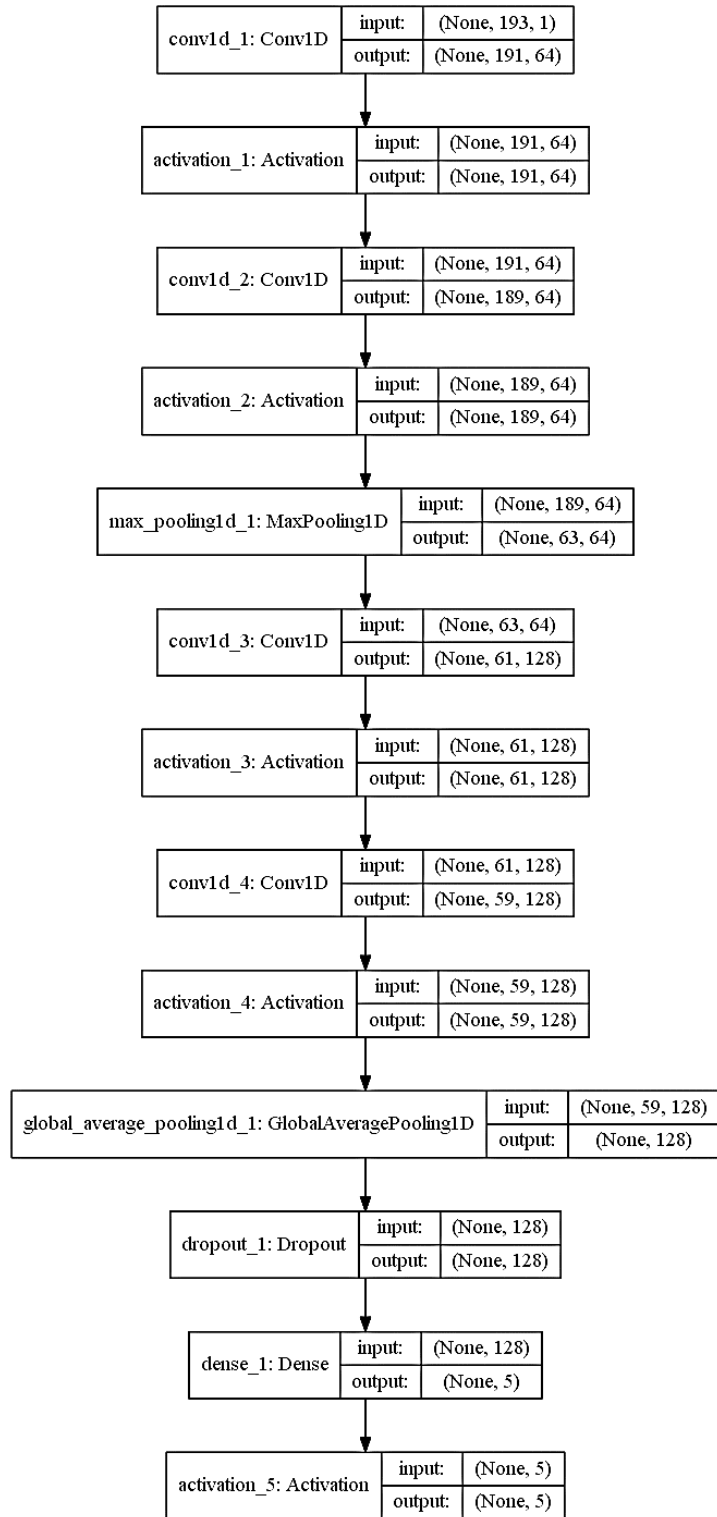
| conv1d_1: Conv1D | input: | (None, 193, 1) |
| | output: | (None, 191, 64) |

| activation_1: Activation | input: | (None, 191, 64) |
| | output: | (None, 191, 64) |

| conv1d_2: Conv1D | input: | (None, 191, 64) |
| | output: | (None, 189, 64) |

| activation_2: Activation | input: | (None, 189, 64) |
| | output: | (None, 189, 64) |

| max_pooling1d_1: MaxPooling1D | input: | (None, 189, 64) |
| | output: | (None, 63, 64) |

| conv1d_3: Conv1D | input: | (None, 63, 64) |
| | output: | (None, 61, 128) |

| activation_3: Activation | input: | (None, 61, 128) |
| | output: | (None, 61, 128) |

| conv1d_4: Conv1D | input: | (None, 61, 128) |
| | output: | (None, 59, 128) |

| activation_4: Activation | input: | (None, 59, 128) |
| | output: | (None, 59, 128) |

| global_average_pooling1d_1: GlobalAveragePooling1D | input: | (None, 59, 128) |
| | output: | (None, 128) |

| dropout_1: Dropout | input: | (None, 128) |
| | output: | (None, 128) |

| dense_1: Dense | input: | (None, 128) |
| | output: | (None, 5) |

| activation_5: Activation | input: | (None, 5) |
| | output: | (None, 5) |

Fig.10. TFCNN network architecture

Table 1. Experimental Results

	Input	Accuracy
VGG16	Spectrogram	**91.46% ±1.02%**
TFCNN	Raw audio signal	**86.33% ± 3.06%**

D. TFCNN Function:

The above table (Table 1) specifies the performance of the model in terms of average accuracy and standard deviation on the test data. As observed, using image spectroscopy, the VGG16 network presents a better performance in classifying fluids sound compared to TFCNN network, which performs classification using the raw audio signal. The reason for the high standard deviation in the TFCNN network performance compared to VGG16 is that some of the signals have an inherent noise caused by the test error and the testing equipment. The amount of this noise varies from data to another, making difficult to eliminate the noise and cause misclassification.

Table 2. Processing information

Processing system	Google collaborator
Processing time	0.56s

3. Conclusions

The present study sought to present two networks for detecting the temperature of a fluid based on its sound, where TFCNN network worked directly with the raw audio signal of the fluid and performed the classification on data. On the other hand, the other network with the VGG16 architecture worked on the spectrogram images of audio signals as inputs and transformed the problem of fluids sound classification into the classification of spectrogram images resulted from the input signal. In order to achieve better performance, a series of preprocessing operations were performed on training data (both audio signal and spectrogram images), which improved the accuracy of the network performance in the classification of fluids sound. The data used in this research were collected through fluid sounds with different temperatures and used in the classification of fluids based on temperature. This paper represents the first research on the relation between sound of the fluids while colliding with a surface and their temperature. By using more advanced equipment such as more efficient heating system and high-level microphone, results would be more accurate. Also, this research could be expanded by studying the relation between sound of other fluids while colliding with a surface and their temperature.

Acknowledgments

References

[1] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological cybernetics. 1980;36(4):193-202.
[2] LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. Neural computation. 1989;1(4):541-51.
[3] Wang W, Lo H-K. Machine Learning for Optimal Parameter Prediction in Quantum Key Distribution. arXiv preprint arXiv:181207724. 2018.
[4] Lu F-Y, Yin Z-Q, Wang C, Cui C-H, Teng J, Wang S, et al. Parameter optimization and real-time calibration of a measurement-device-independent quantum key distribution network based on a back propagation artificial neural network. JOSA B. 2019;36(3):B92-B8.
[5] Nautrup HP, Delfosse N, Dunjko V, Briegel HJ, Friis N. Optimizing quantum error correction codes with reinforcement learning. arXiv preprint arXiv:181208451. 2018.
[6] Beach MJ, De Vlugt I, Golubeva A, Huembeli P, Kulchytskyy B, Luo X, et al. QuCumber: wavefunction reconstruction with neural networks. arXiv preprint arXiv:181209329. 2018.
[7] Ramezanpour A. Enhancing the efficiency of quantum annealing via reinforcement: A path-integral Monte Carlo simulation of the quantum reinforcement algorithm. Physical Review A. 2018;98(6):062309.
[8] Noé F, Wu H. Boltzmann generators-sampling equilibrium states of many-body systems with deep learning. arXiv preprint arXiv:181201729. 2018.

[9] Giannetti C, Lucini B, Vadacchino D. Machine Learning as a universal tool for quantitative investigations of phase transitions. Nuclear Physics B. 2019;944:114639.

[10] Schäfer F, Lörch N. Divergence of predictive model output as indication of phase transitions. arXiv preprint arXiv:181200895. 2018.

[11] Kashiwa K, Kikuchi Y, Tomiya A. Phase transition encoded in neural network. Progress of Theoretical and Experimental Physics. 2019;2019(8).

[12] Quek Y, Fort S, Ng HK. Adaptive Quantum State Tomography with Neural Networks. arXiv preprint arXiv:181206693. 2018.

[13] Nichols R, Mineh L, Rubio J, Matthews JC, Knott PA. Designing quantum experiments with a genetic algorithm. arXiv preprint arXiv:181201032. 2018.

[14] O'Driscoll L, Nichols R, Knott P. A hybrid machine learning algorithm for designing quantum experiments. Quantum Machine Intelligence. 2019;1(1-2):5-15.

[15] Sorteberg WE, Garasto S, Pouplin AS, Cantwell CD, Bharath AA. Approximating the solution to wave propagation using deep neural networks. arXiv preprint arXiv:181201609. 2018.

[16] Ming Y, Lin C-T, Bartlett SD, Zhang W-W. Quantum topology identification with deep neural networks and quantum walks. arXiv preprint arXiv:181112630. 2018.

[17] Varsamopoulos S, Bertels K, Almudever CG. Designing neural network-based decoders for surface codes. arXiv preprint arXiv:181112456. 2018.

[18] Bohrdt A, Chiu CS, Ji G, Xu M, Greif D, Greiner M, et al. Classifying snapshots of the doped hubbard model with machine learning. Nature Physics. 2019:1-4.

[19] Flurin E, Martin LS, Hacohen-Gourgy S, Siddiqi I. Using a recurrent neural network to reconstruct quantum dynamics of a superconducting qubit from physical observations. arXiv preprint arXiv:181112420. 2018.

[20] Goodfellow I, Bengio Y, Courville A. Deep learning: MIT press; 2016.

[21] Wang J, Ma Y, Zhang L, Gao RX, Wu D. Deep learning for smart manufacturing: Methods and applications. Journal of Manufacturing Systems. 2018;48:144-56.

[22] Ng A, editor what data scientists should know about deep learning. Extract data conference; 2017 march: Baidu Research.

[23] Ciaburro G, Venkateswaran B. Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles: Packt Publishing Ltd; 2017.

[24] Gollapudi S. Deep Learning for Computer Vision. Learn Computer Vision Using OpenCV: Springer; 2019. p. 51-69.

[25] Gulli A, Pal S. Deep Learning with Keras: Packt Publishing Ltd; 2017.

[26] Thampi SM, Mitra S, Mukhopadhyay J, Li K-C, James AP, Berretti S. Intelligent Systems Technologies and Applications: Springer; 2018.

[27] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. arXiv preprint arXiv:14042188. 2014.

[28] Howard AG. Some improvements on deep convolutional neural network-based image classification. arXiv preprint arXiv:13125402. 2013.

[29] CireşAn D, Meier U, Masci J, Schmidhuber J. Multi-column deep neural network for traffic sign classification. Neural networks. 2012;32:333-8.

[30] Goodfellow IJ, Bulatov Y, Ibarz J, Arnoud S, Shet V. Multi-digit number recognition from street view imagery using deep convolutional neural networks. arXiv preprint arXiv:13126082. 2013.

[31] Sermanet P, Chintala S, LeCun Y. Convolutional neural networks applied to house numbers digit classification. arXiv preprint arXiv:12043968. 2012.

[32] Men K, Dai J, Li Y. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. Medical physics. 2017;44(12):6377-89.

[33] Ciresan D, Giusti A, Gambardella LM, Schmidhuber J, editors. Deep neural networks segment neuronal membranes in electron microscopy images. Advances in neural information processing systems; 2012.

[34] Lee H, Pham P, Largman Y, Ng AY, editors. Unsupervised feature learning for audio classification using convolutional deep belief networks. Advances in neural information processing systems; 2009.

[35] Deng L, Abdel-Hamid O, Yu D, editors. A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion. ICASSP; 2013.

[36] Abdel-Hamid O, Mohamed A-r, Jiang H, Deng L, Penn G, Yu D. Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and language processing. 2014;22(10):1533-45.

[37] Sainath TN, Mohamed A-r, Kingsbury B, Ramabhadran B, editors. Deep convolutional neural networks for LVCSR. 2013 IEEE international conference on acoustics, speech and signal processing; 2013: IEEE.

[38] Abdel-Hamid O, Mohamed A-r, Jiang H, Penn G, editors. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. 2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP); 2012: IEEE.

[39] Abdel-Hamid O, Deng L, Yu D, editors. Exploring convolutional neural network structures and optimization techniques for speech recognition. Interspeech; 2013.

[40] Deng L, Li J, Huang J-T, Yao K, Yu D, Seide F, et al., editors. Recent advances in deep learning for speech research at Microsoft. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; 2013: IEEE.

[41] Noman F, Ting C-M, Salleh S-H, Ombao H, editors. Short-segment heart sound classification using an ensemble of deep convolutional neural networks. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019: IEEE.

[42] Barchiesi D, Giannoulis D, Stowell D, Plumbley MD. Acoustic scene classification: Classifying environments from the sounds they produce. IEEE Signal Processing Magazine. 2015;32(3):16-34.

[43] Chachada S, Kuo C-CJ. Environmental sound recognition: A survey. APSIPA Transactions on Signal and Information Processing. 2014;3.

[44] Piczak KJ, editor Environmental sound classification with convolutional neural networks. 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP); 2015: IEEE.
[45] Zhu B, Xu K, Wang D, Zhang L, Li B, Peng Y, editors. Environmental sound classification based on multi-temporal resolution convolutional neural network combining with multi-level features. Pacific Rim Conference on Multimedia; 2018: Springer.
[46] Piczak KJ, editor ESC: Dataset for environmental sound classification. Proceedings of the 23rd ACM international conference on Multimedia; 2015: ACM.
[47] Zhu B, Wang C, Liu F, Lei J, Huang Z, Peng Y, et al., editors. Learning environmental sounds with multi-scale convolutional neural network. 2018 International Joint Conference on Neural Networks (IJCNN); 2018: IEEE.
[48] Medhat F, Chesmore D, Robinson J, editors. Masked conditional neural networks for environmental sound classification. International Conference on Innovative Techniques and Applications of Artificial Intelligence; 2017: Springer.
[49] Dieleman S, Brakel P, Schrauwen B, editors. Audio-based music classification with a pretrained convolutional network. 12th International Society for Music Information Retrieval Conference (ISMIR-2011); 2011: University of Miami.
[50] Van den Oord A, Dieleman S, Schrauwen B, editors. Deep content-based music recommendation. Advances in neural information processing systems; 2013.
[51] Medhat F, Chesmore D, Robinson J, editors. Automatic classification of music genre using masked conditional neural networks. 2017 IEEE International Conference on Data Mining (ICDM); 2017: IEEE.
[52] Nasrullah Z, Zhao Y. Music Artist Classification with Convolutional Recurrent Neural Networks. arXiv preprint arXiv:190104555. 2019.
[53] Feng L, Liu S, Yao J. Music genre classification with paralleling recurrent convolutional neural network. arXiv preprint arXiv:171208370. 2017.
[54] Clark S, Park D, Guerard A. Music genre classification using machine learning techniques. Citeseer; 2012.
[55] Shuvaev S, Giaffar H, Koulakov AA. Representations of sound in deep learning of audio features from music. arXiv preprint arXiv:171202898. 2017.
[56] Hussain M, Haque MA. Swishnet: a fast convolutional neural network for speech, music and noise classification and segmentation. arXiv preprint arXiv:181200149. 2018.
[57] Al-Shemmeri T. Engineering fluid mechanics: Bookboon; 2012.
[58] Becker S, Ackermann M, Lapuschkin S, Müller K-R, Samek W. Interpreting and explaining deep neural networks for classification of audio signals. arXiv preprint arXiv:180703418. 2018.
[59] Dahan E, Keller Y. SelfKin: Self Adjusted Deep Model For Kinship Verification. arXiv preprint arXiv:180908493. 2018.
[60] Abadi M, editor TensorFlow: learning functions at scale. Acm Sigplan Notices; 2016: ACM.
[61] Nagi J, Ducatelle F, Di Caro GA, Cireşan D, Meier U, Giusti A, et al., editors. Max-pooling convolutional neural networks for vision-based hand gesture recognition. 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA); 2011: IEEE.
[62] Lin M, Chen Q, Yan S, editors. Network in network. arXiv preprint arXiv: 13124400 2013. Workshop on Statistical Atlases and Computational Models of the Heart Springer; 2011.

Authors' Profiles

ARSHIA FOROOZAN YAZDANI is currently studying in allameh helli college. His articles are mainly about biochemistry and computational physics. He intends to continue his education in biotechnology field.

ALI BOZORGI MEHR is a student in allameh helli college he published his first article at the age of 15. He intends to continue his education in theoretical physics. He is interested in deep learning studies and quantum physics.

IMAN SHOWKATYAN received the B.Sc. degree in physics from Shahid Beheshti University (SBU) and he received his Master in medical physics with image processing tendency in Iran University of Medical Science in 2018 and 2020. He is active researcher in the area of Image Processing, Machine Learning, Deep Learning, Radiomics and computer vision.

AMIN HASHEMI has a background in Physics. He got his Bachelor's degree in solid state physics at Shahid Beheshti University (SBU). He then joined the Laser and Plasma research institute of SBU. Amin finally completed his Master's degree in Plasma Engineering. His thesis is about conversion of toxic gases with plasma into harmless components.

MOHSEN KAKAVAND received the Ph.D. degree in intelligent computing from the University Putra Malaysia (UPM), Malaysia in 2017. He is currently a Senior Lecturer with the Department of Computing and Information Systems, Faculty of Science and Technology, Sunway University in Malaysia. His research interests include aspects of data mining, intelligent computing, machine learning, intrusion detection systems (IDSs), and cybersecurity.