Journal of Theoretical and Applied Information Technology

15th October 2023. Vol.101. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FOREST FIRE DETECTION AND MONITORING THROUGH ENVIRONMENT SOUND SPECTRUM USING DEEP LEARNING

MOUNIR GRARI[1], MOHAMMED BOUKABOUS[1], MIMOUN YANDOUZI[2], MOHAMMED BERRAHAL[1], IDRISS IDRISSI[1], OMAR MOUSSAOUI[1], MOSTAFA AZIZI[1], MIMOUN MOUSSAOUI[1]

[1]MATSI Lab., ESTO, Mohammed First University, Oujda, Morocco

[2]LSI Lab., ENSAO, Mohammed First University, Oujda, Morocco

E-mail: m.grari@ump.ac.ma

ABSTRACT

Forests are one of the most important ecosystems on Earth. They play a vital role in regulating the climate and act as a renewable source of air for human beings. However, forests are really threatened by fires. When wildfires occur outside of their natural range or size, they can become a real danger to life and property. In this paper, we propose an original novel approach for detecting forest wildfires, based on collected data of fire sounds. This method employs a deep learning (DL) model to analyze and classify environment sounds into two classes: “Fire” or “No fire” (usual forest sounds). The model must first be trained on a set of environmental sounds in order to learn and identify fire sound patterns from other sounds. With this model, we achieve an impressive accuracy of 94.24% on the testing sub-dataset. Notably, the model consists of only 1789 parameters, rendering it exceptionally lightweight. This quality makes it highly conducive for deployment across various platforms such as IoT devices, embedded systems, or mobile devices. Integrating this model into forest environments and fortifying it with complementary tools for comprehensive validation could enable us to promptly notify decision-makers or relevant authorities, facilitating timely and decisive actions.

Keywords: *Forest Fires, Wildfires, Deep Learning (DL), Sound spectrum, Audio environment*

1. INTRODUCTION

Forests cover almost one-third of the world's land surface, making them one of the most important ecosystems. They are home to more than two-thirds of the terrestrial biodiversity, including many endangered species. Forests provide critical ecosystem services, including climate change mitigation, carbon sequestration, soil erosion control, and water purification [1]. They are also key elements to the livelihoods of millions of people around the world, providing wood for fuel and construction, food, and traditional medicine. But forests are vulnerable to catastrophes such as wildfires.

Wildfires are a natural part of the forest life cycle. However, when they occur in areas with a high level of human activities, they can cause great damage to both the natural environment and human infrastructure. In addition, wildfires can spread quickly and be difficult to control, making them a serious threat to life and property.

Many factors can contribute to a wildfire, such as a drought, lightning, and careless human activities. In recent years, the number of wildfires has been increasing, due in part to climate change. As the world gets warmer, dry conditions become more common and augment the readiness of forests to be fired.

Wildfires can have a devastating impact on forests. They can destroy trees, wildlife habitats, and watersheds. They can also release pollutants into the air and cause respiratory problems for people. In some cases, wildfires can even lead to the loss of human life [2]. While wildfires can make great harm, they also provide some benefits to the forests; they can help to control insect populations and promote the growth of new plants. When wildfires occur in a natural setting, they can improve the overall health of the forest. However, the negative impacts of wildfires far outweigh the positive ones. That is why it is so important to take steps to prevent them.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Early detection of wildfires is critical to containing these disasters. Most countries use traditional methods for detecting fires including lookouts, ground patrols, and aircraft, others are implementing the latest technologies, which enable us to develop novel methods for detecting and monitoring forest fires in their earliest stages and spread. One common method is using aerial detection, with satellites, aircraft, or drones [3]. Another method is using ground-based IoT sensors, installed accurately in a network inside a forest. These sensors can detect heat, smoke, and other indicators of a fire [4].

Our contribution in this paper lies in a novel approach to detect forest wildfires. This method involves utilizing initial sound data and potentially integrating additional sensors, such as temperature, humidity, carbon monoxide detectors, and cameras. Employing deep learning models, we conduct automated analysis of fire-related acoustic signals, enabling the identification of distinctive patterns associated with fires. This information could be of great usefulness to alert the appropriate authorities to take suitable steps.

The remainder of the paper is organized as follows. The second section gives a background on the utilized techniques. The third section reports and discusses related studies. We present in the fourth section our proposed method. Before concluding, the fifth section reviews and discusses our findings.

2. BACKGROUNDS

2.1 Sensors for Forest Fire

Sensors are devices that measure or detect a physical or chemical property and then convert it into a signal that can be read by an observer or a recording device. There are all sorts of different sensors out there that measure all sorts of different things [1]. Temperatures, pressure, light, sound, electricity, and magnetism are just a few of the many things that sensors can measure. They are also used in scientific research, industrial processes, and manufacturing [5].

One of the most common types of sensors that play an important role in detecting and fighting wildfires are:

- ✓ The temperature sensor, come in all shapes and sizes but they all serve the same purpose: to measure the temperature of something. Temperature sensors are one of the most important types of sensors for detecting wildfires. By measuring the temperature of the air, these sensors can give an early warning of a potential fire.
- ✓ Humidity sensors are another important type of sensor for detecting wildfires. By measuring the humidity of the air, these sensors can help to identify areas at risk of wildfires.
- ✓ Gaz sensors are another type of sensor that can be used for detecting wildfires. By measuring the level of gas in the air, these sensors can give an early warning of a potential fire.
- ✓ Infrared sensors are also another type of sensor that can be used for detecting wildfires. These sensors can detect the heat signature of a fire, even if it is not yet visible to the naked eye.
- ✓ Another sensor is the Microphone; which is a transducer that converts sound into an electrical signal. The converted electrical signal can be used to perform various operations.

2.2 Sound

Sound is a type of energy that travels through the air, or any other medium, as a vibration of pressure waves. Sound waves spread, when something vibrates; the frequency of the wave determines the pitch of the sound. It is the number of times per second a wave oscillates. Hertz (Hz) is the unit used for quantifying the frequency. The human ear can detect sound waves with frequencies ranging from 20Hz to 20 kHz [6].

The loudness of a sound wave is determined by its amplitude; the more pressure waves present, the louder is the sound. The human ear can detect sounds with amplitudes as low as one trillionth of an atmospheric pressure change. The sound's speed is dependent on the medium in which it travels. Sound waves travel the fastest through solids, then liquids, and finally gases; the speed of sound in air is approximately 343 m/s [7],[8].

Fire sounds are one of the most recognizable and distinctive signals in the world. When it comes to fire, the sound of fire is unmistakable [8], whether it is a small campfire or a raging forest fire; the sound of the crackling flames, the hiss of embers, and the snapping of flames, as well as wood popping and exploding, is both soothing yet exhilarating. Burning wood and other bushes produce a crackling sound as the fire consumes them. We also hear a whooshing sound as the air is dragged into the fire. The combination of these two sounds (crackling and whooshing) creates a distinctive sound often associated with fires. When combined with steam produced by the evaporation of moisture out from firewood, the waste gases from combustion enlarge due to the heat and must flee.

The level of noise generated by fire depends on the moisture content of the burning wood, and the type of tree or bushes. Dried wood with a high moisture content will crackle and pop more. When heat expansion forces water vapor to get out from the wood by forming steam. Because of this, the fire erupts with popping and crackling sounds. As a result, when the firewood has a high moisture content, the sound of popping and crackling can be heard more clearly [9].

2.3 Deep Learning (DL)

Deep Learning is a Machine Learning technique that uses a deep neural network to learn from data. Deep Learning is used to solve complex problems that are difficult to solve using traditional machine learning methods [10], [11]. Deep learning is well suited for problems that are difficult to solve using traditional methods, such as image recognition and classification, natural language processing, and machine translation. Deep learning is used in many applications such as computer vision, natural language processing, and predictive analytics [12].

Convolutional neural networks (CNNs) are one of the subclasses of deep learning neural networks that are commonly used to analyze visual data [13]. CNN, like standard neural networks, consists of an input layer, an output layer, and several hidden layers in between, but on the other hand, are distinguished from normal neural networks by their ability to extract features from input images and utilize the spatial relationship between pixels in order to improve pattern recognition [14].

CNN's have been applied to a number of applications, such as picture classification, object identification, and face recognition. CNN has attained state-of-the-art performance on a number of these tasks in recent years [15].

Typically, CNNs are constructed using a succession of convolutional layers followed by fully linked layers, convolutional layers extract information from the input images and forward them to the subsequent layer, and fully connected layers combine the convolutional layers' collected features and utilize them to produce predictions (see Fig. 1) [16].

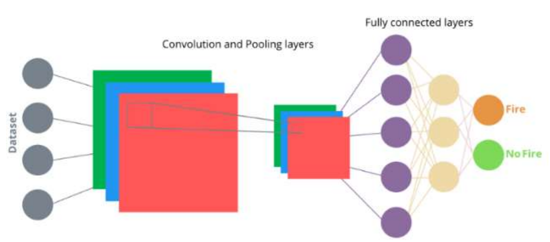

Fig. 1. The architecture of the CNN model

3. RELATED WORKS

Based on IoT, Zhang et al. [8] put forth an original approach to wildfire detection. This method employs sound spectrum analysis to differentiate between crown fires and surface fires. The researchers innovatively designed a tree-energy device, utilizing the inherent biological energy of living trees for generating electricity. The integration of classification algorithms followed the analysis of sound sensor data through this process. According to their findings, the sound frequency of 0–400 Hz signifies crown fires, whereas frequencies of 0–15,000 Hz indicate surface fires. However, the precision of the classification technique is subject to influences like sensor distribution, sound transmission loss, and data transmission delay. In simulated experiments, the method achieved a recognition rate of approximately 70%.

Correia et al. [17] employed energy measurements to track a mobile acoustic source for wildfire detection using drone-based acoustic sensors. This scenario is challenging due to drone motion. They used an extended Kalman filter (EKF) to handle nonlinearities in the observation model, demonstrating its effectiveness in noisy environments and superiority over techniques neglecting prior process state knowledge.

Bai et al. [18] used deep learning to improve audio quality with reverberation captured via multiple microphones. They proposed a method addressing acoustic signal processing: reducing reverberation and categorizing sound events (SEC). The study compared neural networks to traditional techniques like beamforming, multi-channel inverse filtering, multi-channel Wiener filtering, and variance-normalized delayed linear prediction (NDLP) for dereverberation. Their approach showed up to 85% performance improvement in F1-score.

4. PROPOSED METHOD

Our novel approach relies on using IoT sensors for detecting forest wildfires. The focus is oriented to sounds rather than observable flames. These new sound sensors will reinforce other used classical sensors and will be deployed attached to trees in safe positions from both animals and humans while gathering the most data in the air with the maximum coverage. It transmits this information over Lora protocol to the fog gateway, once these sensors have collected data on temperature, humidity, and CO. Ideally, the gateway at the fog node should be positioned in a high location so that it can receive and transmit data from and to IoT devices without interference. It processes the collected data. When it comes to sound data, deep learning will be applied to automatically recognize patterns that may indicate the presence of an active fire in the forest. Using this information, authorities may be alerted and take necessary actions to put out the wildfire. Data afterward is delivered to cloud servers for additional

Journal of Theoretical and Applied Information Technology
15th October 2023. Vol.101. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

analysis, storage, and dashboarding utilizing lightweight MQTT protocol via 2/3/4/5G, satellite internet, or any possible communication technique available at the forest location (see Fig. 2).

Fig. 2. Proposed method deployment of sound sonsors

We conduct our research experiment in three phases: dataset collecting, spectrogram production, and finally, the construction of a sound deep learning model (see Fig. 3).

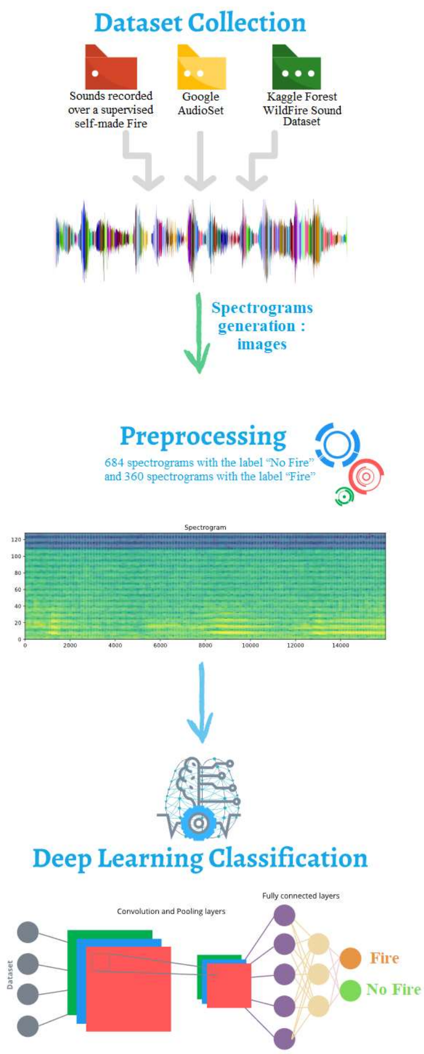

Fig. 3. Research experiment phases

4.1 Sound Dataset collection

Datasets of sounds can be collected in various ways, depending on the desired application [11]. For example, a speech recognition system might be trained on a dataset of recorded speech samples, while a system for identifying environmental sounds might be trained on a dataset of natural sounds like raindrops or birdcalls. In general, collecting a good dataset of sounds usually involves recording a variety of sounds in different environments and at various times [19]. We make this task with special-purpose sound recording equipment, or with a pair of regular microphones and a digital audio recorder.

In our experiments, the dataset used to train our models is derived from three sources: Google AudioSet (a vast collection of human-labeled ten seconds environmental sound snippets obtained from YouTube), A Kaggle Forest WildFire Sound Dataset, and a set of sounds recorded with our smartphone microphone over a supervised self-made fire with local tree wood (see Fig. 4). The collected sounds are divided into two categories: the first one is named "Forest_environment", it contains the major sounds that may be heard in a forest, coming from Animals, Birds, Insect, Rain, Waterfall, Wind, Speech, Silence, Vehicles, and Aircraft; the second category includes only Fire sounds. What results, after cleaning mixed and corrupted sounds, into 1044 samples in total; 684 with the label "Forest environment" which correspond to the "No Fire" output of our DL model, and 360 with the label "Fire". Afterward, we divided the dataset into training (60%), validation (20%), and testing (20%) sets, before any further preprocessing or training.

Fig. 4. Supervised autonomous fire sound recording

The amplitude of the waveform is shown on the y-axis of a time series of these audio samples (see Fig. 5). It is standard practice to quantify amplitude in terms of the change in pressure around the microphone or receiving device that first detected the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

sound. In order to train our model, we used these time series signals as inputs. A visual analysis of the samples below (see Fig. 5) taken from each of the two classes in our dataset (Fire and Environment sounds) reveals that the waveform does not give obvious class identification information.

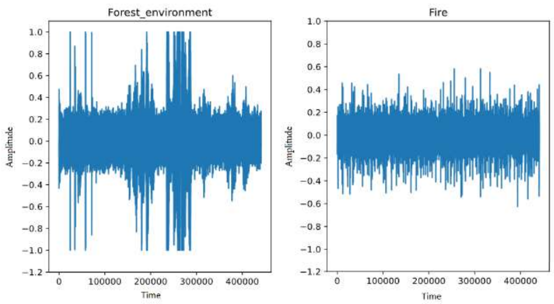

Fig. 5. Sound waveforms of the two categories "Fire" and "Forest environment"

4.2 Spectrograms generation

To perform the generation of spectrograms from temporal waveforms, we apply the short-time Fourier transform (STFT) [20]. STFT allows obtaining time-localized frequency information, while frequency components of a signal fluctuate over time and it may be so complex to make conclusions. The typical Fourier transform provides frequency information averaged across the full signal time period. STFTs are given by formula 1 [21]. As shown in Fig. 5, STFT turns the waveforms into spectrograms. These latter are a kind of two-dimensional imagery representations of a signal frequency spectrum. Our dataset will be a collection of sound spectrograms. We need these spectrogram images to train our DL model [22].

$$\begin{cases} \mathrm{X_{STFT}}[m,n] = \sum_{\mathrm{k=0}}^{\mathrm{L-1}} x[k]g[k-m]\mathrm{e}^{\frac{-\mathrm{j}2\pi \mathrm{n}}{\mathrm{L}}} \\ \mathrm{x[k]} = \sum\sum \mathrm{X_{STFT}}[m,n]\, g[k-m]\mathrm{e}^{\frac{\mathrm{j}2\pi \mathrm{nk}}{\mathrm{L}}} \end{cases} \quad (1)$$

where $x[k]$ represents a signal and $g[k]$ characterizes an L-point window, the STFT of $x[k]$ can be viewed as the Fourier transform of the product $x[k]g[k-m]$.

Each frequency's amplitude or intensity is illustrated by a distinct color, (The more vibrant the hue, the more powerful the message). The Spectrogram's vertical 'slices' represent the signal's Spectrum at that point in time, and they display the intensity of the signal at each frequency in that spectrum.

For example, in the samples below (see Fig. 6), the first top images show amplitude versus time. However, it does not provide us with any information on which frequencies are present in a particular piece of sound at any given moment. The bottom images are spectrograms, which represent the signal in its frequency domain.

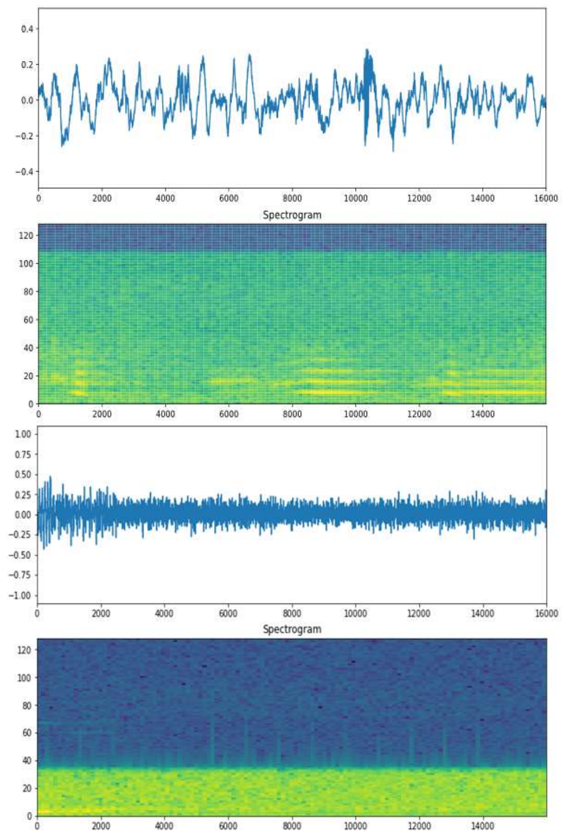

Fig. 6. Examples of generated sound spectrograms from waveforms using STFT

4.3 Our Sound Deep Learning Model

Our proposed deep learning model is a neuronal network composed of a sequence of multiple layers (see **Error! Reference source not found.**) [23]:

- ✓ The input layer that resizes the input images to 64x64;
- ✓ The second layer normalizes the images. It is typically used to ensure that the activations of the neurons in the network stay within a certain range, which makes training more stable and improves the generalization performance of the model [24];
- ✓ The convolution 2d layer convolves the images with 32 filters. A convolution is an operation that takes two inputs, a filter, and an input image, and produces an output image. The output image is produced by the convolution operation, which is a matrix multiplication between the input image and the filter [16];

Journal of Theoretical and Applied Information Technology
15th October 2023. Vol.101. No 19

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

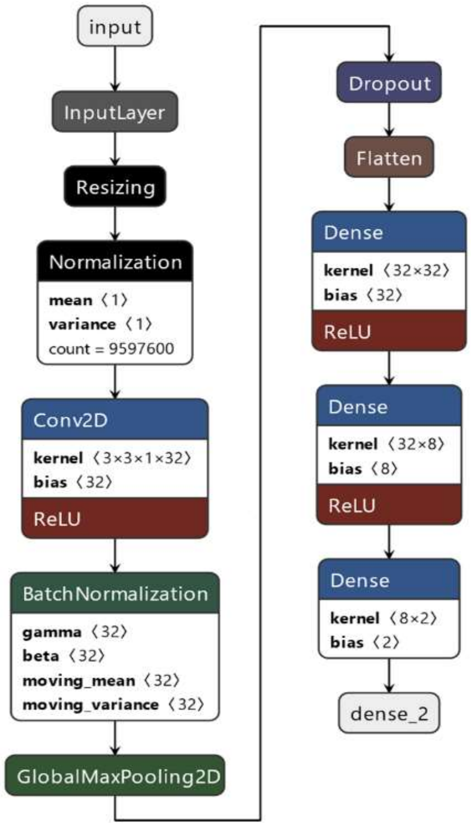

Fig. 7. Our proposed model architecture

- ✓ The batch normalization layer batch-normalizes the convolved images. Batch normalization is typically used after the convolutional or fully connected layers, while layer normalization can be used anywhere in the network. Both batch normalization and normalization layers have been shown to be very effective at improving the training and generalization performance of deep neural networks. In many cases, they can even allow the use of much higher learning rates, which can further speed up training [25], [26]. The global max pooling 2d layer pools the maximum value from each of the 32 filters [3]. It is similar to the max pooling2d layer, but it computes the maximum overall channels for each spatial position, instead of computing the maximum over a window for each channel. This layer is useful for tasks such as image classification, where the input is an image with multiple channels (e.g., a 3-channel RGB image). By computing the maximum overall channels, this layer can create a summary of the image that is invariant to the input channel order.
- ✓ The dropout layer drops out a random 20% of the neurons. This layer is a regularization technique for neural networks that helps prevent overfitting. The dropout layer randomly drops out (sets to zero) a number of output units from the previous layer during training. The dropped-out values are chosen randomly [27].
- ✓ The flatten layer that flattens the 32 filters into a vector. A flattened layer takes an input with a potentially high-dimensionality (such as a 2D tensor with the shape: samples, channels, rows, cols) and flattens it into a 1D tensor with shape (samples, flattened dimension). This is useful for feeding dense layers since a dense layer expects its input to be 1D [28].
- ✓ The three dense layers (fully connected layers) that have respectively 1056, 264, 18 neurons.
- ✓ The output layer has 2 neurons (for the Fire or environmental sounds classes).

The total number of parameters is 1789, which means that our proposed model is a lightweight one (because it is relatively small and does not require a lot of computational resources compared to other deep learning models), and would be perfect for use in a resource-constrained environment such as IoT devices, mobile devices, or embedded systems. This model would be able to run quickly and efficiently while still providing good accuracy.

5. RESULTS AND DISCUSSIONS

5.1 Hardware and software characteristics

For our implementation, we have used TensorFlow v2.8.0, an open-source data analysis and machine learning software library, on a high-performance computing system (HPC) equipped with the following hardware specifications:

- ✓ Two Intel Gold 6148 (2.4 GHz/20 cores) processors;
- ✓ Two NVIDIA Tesla V100 graphics cards, each with 32GB of RAM.

5.2 Evaluation Metrics

We evaluated our deep learning models using two metrics namely, Sparse Categorical Cross-Entropy Loss, and Accuracy.

5.2.1 Sparse Categorical Cross-Entropy Loss (SCCEL)

Sparse categorical cross-entropy loss is a loss function used in machine and deep learning.

This function is used when the labels are not one-hot encoded. SCCEL allows using integers as labels instead of one-hot encoding them. It is more efficient than categorical cross-entropy loss and is typically used when the number of classes is large [29]. SCCEL is defined by formula 2:

$$\mathrm{SCCEL} = -\sum_{i}^{c} t_i \log(s_i) \qquad (2)$$

where t_i and s_i denote the ground truth and CNN score for each class i in C.

There are several advantages that SCCEL has over the Categorical Cross-Entropy Loss [30]. One advantage is that **SCCEL** is less computationally expensive. This is because **SCCEL** can take advantage of sparsity in the data, which means that it does not need to compute the gradient for every data point. This can be a significant advantage when working with large datasets. Another advantage of **SCCEL** is that it is more stable, this is because it does not suffer from the issues of numerical instability that can occur with Categorical Cross-Entropy Loss.

5.2.2 Accuracy

In predictive modeling, machine, and deep learning, the accuracy metric is used to assess the quality of predictions. The accuracy metric is a measure of how accurate a model is in its predictions. It is the proportion of accurate forecasts to total predictions. This metric could be used to compare various models and identify the most accurate one [31].

5.3 Evaluating the results

We train our proposed model over 300 epochs to monitor its behavior and obtain the best results. The model was inspired by Google's YAMNet, but with several tweaks due to overfitting on our dataset, particularly at batch normalization and global max pooling 2d layers.

Table 1 summarizes the obtained results regarding the accuracy and loss of our proposed model during training, validation, and testing. These results show that the training dataset (respectively validation and test datasets) had an accuracy of 92.17% (respectively 93.69% and 94.24%) and a loss of 23.49% (respectively 23.49% and 18.14%).

Table 1: The obtained results of our model

Dataset	**Metric**	**Results (%)**
Training set	Accuracy	92.17
	Loss	23.49
Validation set	Accuracy	93.69
	Loss	19.62
Testing set	Accuracy	94.24
	Loss	18.14

The model seems to be performing quite well, with increasing accuracy and decreasing loss as the model is trained on much more data (see Fig. 8 and Fig. 9). However, it is important to note that the validation and testing results are very close to each other. This means that the model is not overfitting, and is instead generalizing well to new data. This is a very important property for a deep learning model and is a good indication that this model is performing well.

Fig. 10 displays an example of a sound prediction for fire noises compared to other environmental sounds. We can see that it predicts the fire sounds rather well with certain environmental sounds, which is expected since the microphone sensor detects both sounds.

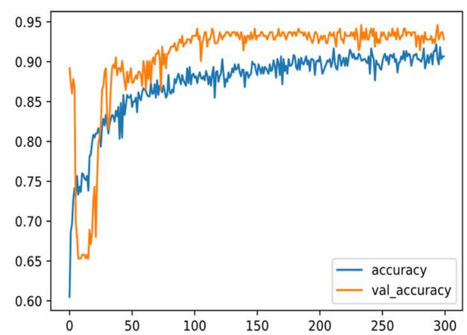

Fig. 8. Accuracy and Validation accuracy curves

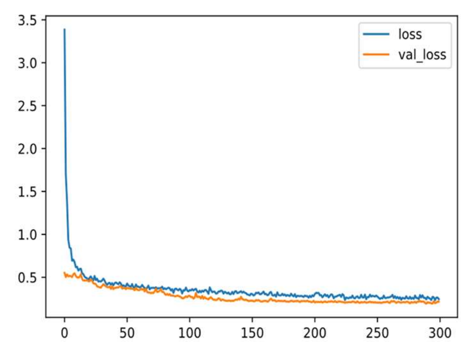

Fig. 9. Loss and Validation loss curves

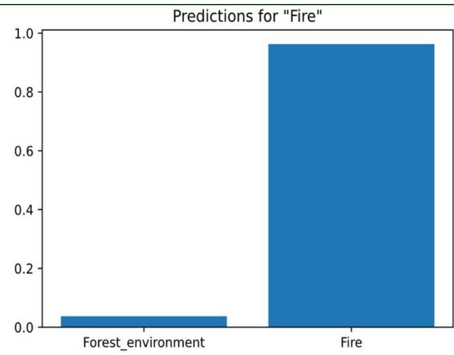

Fig. 10. Sound prediction example

Forest fire detection using sound sensors and deep learning is still in its early stages of development and thus has limitations (the use of sensors in general for forest fire detection has several limitations); sensors are not able to detect all types of fires, including small, contained fires, or smoldering fires (which produce little to no smoke). Meanwhile, our system may be reinforced with other commonly used sensors, such as temperature, humidity, and smoke sensors, which can give supplementary data to increase the precision of forest fire detection. We may further reduce the rate of false positives by dispatching drones to the area around the suspected fire (the location can be acquired from the glocalization of the deployed sensors, such as GPS or Galileo) to determine if a fire is started or not.

6. CONCLUSION

In this study, we introduce a pioneering approach that capitalizes on sound-to-image transformation and leverages the advancements in deep learning within computer vision. Our innovative method focuses on early forest fire detection, eschewing traditional reliance on visible flames in favor of interpreting sound data. By harnessing deep learning, we automatically analyze fire-related acoustic signals, identifying distinctive patterns indicative of a fire's presence, and promptly alerting authorities for necessary actions. Notably, our proposed model achieves a robust 94.24% accuracy on the test dataset while remaining highly efficient with a mere 1789 parameters, making it exceptionally suitable for resource-constrained environments such as IoT and mobile devices.

There are some limitations to this study that will be addressed in our future works, the most notable one is the small size of our used dataset. This dataset contains only 1044 sounds, which is relatively smaller than our goal. Despite its limitations, the results of this study show that the proposed model is a promising approach for the early detection of forest fires. As a perspective, we plan to deploy our proposal inside a forest located in the neighborhood of our university.

ACKNOWLEDGEMENTS

This work is supported by the Mohammed First University. under the PARA1 Program (Low-cost and real-time Forest Fire Detection System based on Wireless Sensor Networks - SDF-RCSF). The used computational resources of HPC-MARWAN are provided by the National Center for Scientific and Technical Research (CNRST). Rabat. Morocco.

REFERENCES:

[1] M. Grari et al., "Using IoT and ML for Forest Fire Detection, Monitoring, and Prediction: a Literature Review," J. Theor. Appl. Inf. Technol., vol. 100, pp. 5445–5461, 2022.

[2] M. Yandouzi et al., "Review on forest fires detection and prediction using deep learning and drones," J. Theor. Appl. Inf. Technol., vol. 100, no. 12, pp. 4565–4576, 2022.

[3] M. Yandouzi et al., "A Lightweight Deep Learning Model for Forest Fires Detection and Monitoring," pp. 697–705, 2023, doi: 10.1007/978-981-19-6223-3_71.

[4] M. Grari, I. Idrissi, M. Boukabous, O. Moussaoui, M. Azizi, and M. Moussaoui, "Early Wildfire Detection using Machine Learning Model Deployed in the Fog / Edge Layers of IoT," Indones. J. Electr. Eng. Comput. Sci., 2022.

[5] I. Idrissi, M. Mostafa Azizi, and O. Moussaoui, "A Lightweight Optimized Deep Learning-based Host-Intrusion Detection System Deployed on the Edge for IoT," Int. J. Comput. Digit. Syst., vol. 11, no. 1, pp. 209–216, 2022, doi: 10.12785/ijcds/110117.

[6] S. A. Fausti, D. A. Erickson, R. H. Frey, B. Z. Rappaport, and M. A. Schechter, "The effects of noise upon human hearing sensitivity from 8000 to 20 000 Hz," J. Acoust. Soc. Am., vol. 69, no. 5, p. 1343, Aug. 1998, doi: 10.1121/1.385805.

[7] "The Soundry: The Physics of Sound." https://www.schoolnet.org.za/PILAfrica/en/webs/19537/physics3.html (accessed Apr. 26, 2023).

[8] S. Zhang, D. Gao, H. Lin, and Q. Sun, "Wildfire Detection Using Sound Spectrum Analysis Based on the Internet of Things," Sensors 2019, Vol. 19, Page 5093, vol. 19, no. 23, p. 5093, Nov. 2019, doi: 10.3390/S19235093.

[9] "Flame noise: sound emission from spark-ignited bubbles of combustible gas," Proc. R. Soc. London. Ser. A. Math. Phys. Sci., vol. 294,

no. 1439, pp. 449–466, Oct. 1966, doi: 10.1098/RSPA.1966.0218.

[10] A. Kherraki and R. El Ouazzani, "Deep convolutional neural networks architecture for an efficient emergency vehicle classification in real-time traffic monitoring," IAES Int. J. Artif. Intell., vol. 11, no. 1, pp. 110–120, Mar. 2022.

[11] I. Idrissi, M. Azizi, and O. Moussaoui, "Accelerating the update of a DL-based IDS for IoT using deep transfer learning," Indones. J. Electr. Eng. Comput. Sci., vol. 23, no. 2, pp. 1059–1067, Aug. 2021, doi: 10.11591/ijeecs.v23.i2.pp1059-1067.

[12] M. Boukabous and M. Azizi, "A comparative study of deep learning based language representation learning models," Indones. J. Electr. Eng. Comput. Sci., vol. 22, no. 2, pp. 1032–1040, 2021, doi: 10.11591/ijeecs.v22.i2.pp1032-1040.

[13] M. Yandouzi et al., "Forest Fires Detection using Deep Transfer Learning," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 8, pp. 268–275, Oct. 2022, doi: 10.14569/IJACSA.2022.0130832.

[14] M. Berrahal and M. Azizi, "Augmented Binary Multi-Labeled CNN for Practical Facial Attribute Classification," Indones. J. Electr. Eng. Comput. Sci., vol. 23, no. 2, pp. 973–979, Aug. 2021.

[15] M. Yandouzi et al., "Investigation of Combining Deep Learning Object Recognition with Drones for Forest Fire Detection and Monitoring," Artic. Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 3, p. 2023, 2023, doi: 10.14569/IJACSA.2023.0140342.

[16] I. Idrissi, M. Boukabous, M. Azizi, O. Moussaoui, and H. El Fadili, "Toward a deep learning-based intrusion detection system for IoT against botnet attacks," IAES Int. J. Artif. Intell., vol. 10, no. 1, pp. 110–120, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp110-120.

[17] S. D. Correia, J. Fé, S. Tomic, and M. Beko, "Drones as Sound Sensors for Energy-Based Acoustic Tracking on Wildfire Environments," pp. 109–125, Nov. 2021, doi: 10.1007/978-3-030-96466-5_8.

[18] M. R. Bai and W. Chen, "Deep learning applied to dereverberation and sound event classification in reverberant environments," in Proceedings of the International Congress on Acoustics, 2019, vol. 2019-Septe, pp. 4917–4923, doi: 10.18154/RWTH-CONV-238800.

[19] M. Boukabous and M. Azizi, "Crime prediction using a hybrid sentiment analysis approach based on the bidirectional encoder representations from transformers," Indones. J. Electr. Eng. Comput. Sci., vol. 25, no. 2, Feb. 2022, doi: 10.11591/IJEECS.V25.I2.PP.

[20] L. Durak and O. Arikan, "Short-time Fourier transform: Two fundamental properties and an optimal implementation," IEEE Trans. Signal Process., vol. 51, no. 5, pp. 1231–1242, 2003, doi: 10.1109/TSP.2003.810293.

[21] N. Kehtarnavaz, "Frequency Domain Processing," Digit. Signal Process. Syst. Des., pp. 175–196, 2008, doi: 10.1016/B978-0-12-374490-6.00007-6.

[22] M. Boukabous and M. Azizi, "Multimodal Sentiment Analysis using Audio and Text for Crime Detection," in 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Mar. 2022, pp. 1–5, doi: 10.1109/IRASET52964.2022.9738175.

[23] I. Idrissi, M. Azizi, and O. Moussaoui, "A Stratified IoT Deep Learning based Intrusion Detection System," in 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Mar. 2022, pp. 1–8, doi: 10.1109/IRASET52964.2022.9738045.

[24] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer Normalization," Jul. 2016, doi: 10.48550/arxiv.1607.06450.

[25] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in 32nd International Conference on Machine Learning, ICML 2015, 2015, vol. 1, pp. 448–456.

[26] N. Bjorck, C. P. Gomes, B. Selman, and K. Q. Weinberger, "Understanding Batch Normalization," in Advances in Neural Information Processing Systems, 2018, vol. 31.

[27] P. Baldi and P. J. Sadowski, "Understanding Dropout," in Advances in Neural Information Processing Systems, 2013, vol. 26.

[28] E. Jeczmionek and P. A. Kowalski, "Flattening Layer Pruning in Convolutional Neural Networks," Symmetry 2021, Vol. 13, Page 1147, vol. 13, no. 7, p. 1147, Jun. 2021, doi: 10.3390/SYM13071147.

[29] "machine-learning-articles/how-to-use-sparse-categorical-crossentropy-in-keras.md at main · christianversloot/machine-learning-articles · GitHub." https://github.com/christianversloot/machine-learning-articles/blob/main/how-to-use-sparse-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

categorical-crossentropy-in-keras.md (accessed Apr. 26, 2023).

[30] D. Ali, M. M. S. Missen, and M. Husnain, "Multiclass Event Classification from Text," Sci. Program., vol. 2021, 2021, doi: 10.1155/2021/6660651.

[31] Y. Hammoudi, I. Idrissi, M. Boukabous, Y. Zerguit, and H. Bouali, "Review on maintenance of photovoltaic systems based on deep learning and internet of things," Indones. J. Electr. Eng. Comput. Sci., vol. 26, no. 2, May 2022.