

# Development of a Framework for Classification of Impulsive Urban Sounds using BiLSTM Network

Nazbek Katayev<sup>1</sup>, Aigerim Altayeva<sup>2</sup>, Bayan Abduraimova<sup>3</sup>, Nurgul Kurmanbekkyzy<sup>4</sup>,  
Zhumabay Madibaiuly<sup>5</sup>, Bakhytzhhan Kulambayev<sup>6</sup>

Kazakh National Women's Teacher Training University, Almaty, Kazakhstan<sup>1</sup>

Al-Farabi Kazakh National University, Almaty, Kazakhstan<sup>2</sup>

L. N. Gumilyov Eurasian National University, Astana, Kazakhstan<sup>3</sup>

Kazakh-Russian Medical University, Almaty, Kazakhstan<sup>4</sup>

Joldasbekov Institute of Mechanics and Engineering, Almaty, Kazakhstan<sup>5</sup>

Academy of Logistics and Transport, Almaty, Kazakhstan<sup>5</sup>

Turan University, Almaty, Kazakhstan<sup>6</sup>

**Abstract**—Urban environments are awash with myriad sounds, among which impulsive noises stand distinct due to their brief and often disruptive nature. As cities evolve and expand, the accurate classification and management of these impulsive sounds become paramount for urban planners, environmental scientists, and public health advocates. This paper introduces a novel framework leveraging the Bidirectional Long Short-Term Memory (BiLSTM) Network for the systematic categorization of impulsive urban sounds. Traditional methodologies often falter in recognizing the nuanced intricacies of such noises. In contrast, the presented BiLSTM-based approach adapts to the temporal variability intrinsic to these sounds, thereby enhancing classification accuracy. The research harnesses an expansive dataset, curated from various urban settings, to train and validate the model. Preliminary findings suggest that our BiLSTM framework outperforms existing models, with a marked increase in both specificity and sensitivity metrics. The outcome of this study holds profound implications for city acoustics management, noise pollution control, and urban health interventions. Moreover, the framework's adaptability paves the way for its application across diverse acoustic landscapes beyond the urban realm. Future endeavors should seek to further optimize the model by integrating more diverse soundscapes and addressing potential biases in data collection.

**Keywords**—Impulsive sound; machine learning; deep learning; CNN; LSTM; classification

## I. INTRODUCTION

In the vibrant tapestry of urban life, sounds and noises play an integral role, shaping the auditory landscapes that city inhabitants navigate daily. The urban soundscape, a combination of ambient noises, human interactions, vehicular movements, and sudden, impulsive sounds, constitutes an integral aspect of urban living [1]. These sounds, particularly the impulsive varieties, serve as a double-edged sword [2]. On one hand, they contribute to the character and ambiance of a city, often evoking deep-seated memories and emotional responses among its residents. On the other hand, unchecked and discordant impulsive noises can deteriorate the quality of life, leading to stress, sleep disturbances, and even chronic health issues [3]. Consequently, the significance of identifying,

classifying, and managing these sounds in urban spaces cannot be overstated.

While a plethora of research has focused on the broad soundscape of cities, the niche area of impulsive urban sounds has traditionally been underserved. Defined by their short, abrupt nature, these sounds—be it the honk of a car, the clang of a dropped tool, or the burst of fireworks—pose unique challenges to classification systems [4]. Traditional audio classification models, built primarily for longer and more consistent sounds, often struggle to capture the fleeting nuances of impulsive noises [5]. The rapid onset and offset of these sounds, combined with their varied frequency range, demand an approach that is both sensitive to temporal dynamics and adaptable to a broad acoustic spectrum.

Enter the realm of neural networks, which in recent years, has revolutionized the domain of sound classification. Among neural architectures, the Long Short-Term Memory (LSTM) network [6], a type of recurrent neural network, has shown promise in handling sequences and time-series data, making it a suitable contender for our auditory challenge. However, a unidirectional LSTM processes data in its input sequence order, potentially overlooking patterns that emerge from the reverse sequence of sounds [7]. Recognizing this limitation, and drawing inspiration from the bidirectional nature of human auditory processing where sounds are often understood in the context of both preceding and following sounds, this research innovatively employs the Bidirectional Long Short-Term Memory (BiLSTM) Network [8]. The BiLSTM, by virtue of processing an input sequence in both forward and backward directions, stands poised to capture the intricate patterns and characteristics intrinsic to impulsive urban sounds, offering a comprehensive understanding of their structure.

This paper, therefore, sets forth with a dual agenda. Firstly, it seeks to elucidate the significance and complexity of impulsive urban sounds, grounding its arguments in both auditory science and urban studies. Secondly, it embarks on a journey to explore the efficacy of the BiLSTM framework in classifying these sounds, aiming to bridge the gap between neural network research and urban acoustic management. In doing so, this research not only endeavors to advance the field

of auditory classification but also aspires to have a tangible impact on urban planning, noise pollution control measures, and public health interventions.

As we delve deeper into this exploration, it becomes imperative to understand the broader context within which urban sounds exist, the technological advancements in neural networks, and the potential applications of an effective classification system. Through this multi-faceted lens, this research hopes to offer a comprehensive view of the challenges and opportunities that lie at the intersection of urban acoustics and advanced neural architectures.

## II. RELATED WORKS

Amid the bustling panorama of urban existence, the cacophony of sounds emerges not just as an incidental backdrop, but as an active participant shaping the dynamics of city life. The auditory fabric of urban centers is woven with diverse threads, ranging from the rhythmic footfalls on pavements to the occasional discordant blare of car horns [9]. Within this intricate web, impulsive urban sounds—transient, unexpected, and often sharp in nature—hold a unique place [10]. Their fleeting existence and unpredictable onset present both an auditory intrigue and a challenge that merit academic and scientific exploration.

As cityscapes continue to evolve, converging towards a future that's increasingly urbanized, the sonic environment they foster becomes an indispensable area of study. The implications of these sounds stretch across various dimensions: psychological, sociological, environmental, and even physiological [11]. For instance, while a distant church bell or a street performer's melody might evoke feelings of nostalgia or joy, the sudden screech of brakes or a loud explosion can trigger stress or anxiety [12]. The dichotomy of these reactions underscores the relevance of understanding and classifying urban sounds, especially those of an impulsive nature [13]. Given their impact on the well-being of city dwellers, mental health, and the broader urban experience, a systematic study becomes not just an academic endeavor but a societal imperative.

Historically, the academic arena has demonstrated a sustained interest in urban noises, resulting in extensive literature on the general soundscape of cities [14]. However, when it comes to the niche area of impulsive sounds, the scholarly attention seems somewhat disproportionate [15]. This relative dearth is surprising, given that impulsive noises, by virtue of their sudden onset and varied frequency profiles, pose unique challenges. Traditional auditory classification models, designed with an inclination towards consistent and prolonged sounds, falter when faced with the erratic nature of impulsive noises [16]. The fleeting presence and diverse acoustic characteristics of these sounds necessitate an approach that's not only nimble but also adept at capturing rapid temporal fluctuations [17].

Enter the world of advanced neural networks—a domain that has, in recent times, transformed numerous fields, including audio processing [18]. Among the neural architectures on offer, the Long Short-Term Memory (LSTM) network, a subtype of recurrent neural networks, has emerged

as a front-runner for tasks involving sequence or time-series data [19]. Given its prowess in handling sequential data, LSTM offers a glimmer of hope for the impulsive sound conundrum. However, traditional LSTM, being unidirectional, processes sequences in the order they are presented, potentially missing out on valuable insights that could be gleaned from reverse order processing.

It's against this backdrop that this research introduces the Bidirectional Long Short-Term Memory (BiLSTM) [20] Network to the equation. Drawing parallels from human auditory processing, which inherently understands sounds based on both their preceding and succeeding context, the BiLSTM processes sequences bidirectionally—both forwards and backwards. This bidirectional approach promises a more holistic grasp of impulsive urban sounds, capturing nuances that might escape unidirectional models [21]. By processing sounds in this dual manner, the BiLSTM aspires to straddle the intricate patterns and temporal dynamics intrinsic to impulsive noises.

This paper embarks on a journey with twofold objectives. First, it aims to contextualize the importance and intricacies of impulsive urban sounds within the broader discourse of urban studies and auditory science [22]. It strives to illustrate why these sounds, often sidelined in scholarly pursuits, deserve focused attention. Second, the research delves into the technical and empirical exploration of the BiLSTM framework, investigating its potential as the much-needed solution to the challenges posed by impulsive sounds. Through this synthesis, the paper hopes to create a bridge—linking the often disparate worlds of neural network research and urban acoustic management.

As we venture further into this academic exploration, we're invited to reflect upon a myriad of interconnected themes: the transformative power of neural networks, the complex tapestry of urban soundscapes, and the potential societal ramifications of effective sound classification. Through this kaleidoscopic lens, this research endeavors to provide a comprehensive and nuanced perspective, setting the stage for groundbreaking revelations in the crossroads of urban acoustics and neural network technology.

## III. MATERIALS AND METHODS

In line with our initial conceptual framework, there is a two-fold requirement: first, to register the designated sound analysis apparatus, and second, to subject it to rigorous training [23]. This machine, once operational, deciphers the ingested auditory data, which can span a gamut of audio formats, a notable example being the mp3 format. Upon the reception of such an audio file, the machine subsequently crafts its associated spectrogram. A spectrogram, often interchangeably termed a sonogram (Fig. 1), is a graphical rendering that elucidates the relationship between the spectral density of a signal's power and its temporal progression [24]. Historically, spectrograms have found multifaceted applications across disciplines. They play a pivotal role in areas such as speech recognition, analysis of animal vocal patterns, diverse musical domains, radio and sonar technologies, linguistic signal processing, seismological research, and several other specialized fields.

### A. Searching and Selecting Dataset

Each model training task requires a lot of input data, and the quality of our model will essentially depend on them. Therefore, the choice of dataset is an important part when building a model. Often the data also needs to be filtered or "cleaned up" in case some of the samples contain misrepresentations or false sounds for the class.

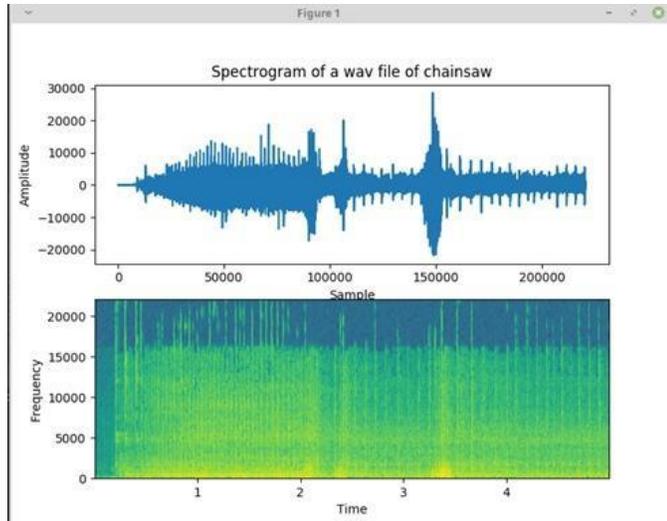


Fig. 1. Example of spectrogram of chainsaw.

We found two very interesting datasets, the first is the UrbanSounds-8K and second is UrbanAudioDataset. But firstly, let's see the first one.

UrbanSounds-8K. This dataset [25] contains about ~900 ".wav" sound files for each 10 classes such as:

- Air conditioners
- Car horns
- Children playing
- Dog barking
- Drilling
- Engine idling
- Gun shots
- Jackhammers
- Sirens
- Street music

Total 8732 audio files. This dataset, is pretty big (about 6,6 gigabytes) but as we know, the more data, the more we can train the model. That's why we believe that it will perfectly show the advantages and disadvantages of model preprocessing. And in the first part of our research we will use it.

Urban Audio Dataset. The second dataset [26] was collected from various resources and consist about 10000 samples for next eight classes:

- Crying
- Dog barking
- Emergency alarms
- Explosions
- Fire
- Glass breaking
- Screaming
- The sound of a weapon firing(gun shots)

However, the data in it requires normalization, since the data format is extremely different (with formats like: .mp3, .aiff, .flac, .wav, .m4a), also weights about 30 gigabytes! Therefore, we will use this dataset only after checking the main model. And this dataset is more suitable for our problem, since these sounds are more suitable for alerting danger. But again, we will talk about this dataset in more detail later in the next parts of the research.

### B. Environment Selection

After the datasets, let's think about the hardware environment. Initially, the development was carried out on a virtual machine in the VirtualBox image on the Linux Mint system, which was sharpened for computer vision tasks, and in particular OpenCV. It fit the prototype, but due to the limitations of virtualization, it was decided to transfer the project to the main machine (host machine) on the Windows 10 operating system.

Now let's decide for the environment itself. Machine learning and deep learning in general are very widely used in the Python language due to ease of use, however, it is worth mentioning that for performance, you can try developing in C / C ++ if the issue of performance will play a key role. But we believe that Python 3.7 is enough for this task.

For analysis and research, we will use Jupyter Notebook. It is on it that we can write lines of code that we can interpret in different ways and thereby observe changes in individual cells of the program launch.

For the final part of development in production, we will be using PyCharm by JetBrains (Community Edition) and IntelliJ IDEA with Java Spring Framework for backend. But about interacting with application we will explain later in next parts.

### C. Sound Processing

In this section, we will talk about the part about digital sounds. We decided that it was very important to fully understand how a computer can pick up sound, how we can work with it, and how to adapt it for classification. In this subsection, we will try to answer these questions.

Digital audio is the result of converting an analog audio signal into a digital audio format. There are a lot of audio formats at the moment, such as .ogg, .wav, .mp3, .flac and more. They differ in their storage and playback properties, but all alone cope with their task - they transmit an audio signal within a digital system. In our case, the sound will be displayed as a graph like in Fig. 2:

From which we can later obtain values after sampling when converting from an analog signal to digital. In other words, the digital sound that is already in the computer is already converted and we can work directly with it.

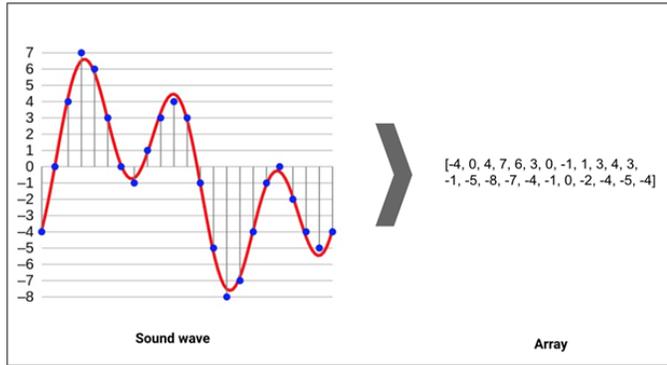


Fig. 2. Example of converting sound wave to array type.

#### D. Sound Analysis and Visualization

Audio has its own sound transmission channels, these channels called “Left-Right” output [27], in simple words, the output of the whole sound goes immediately to both the left ear and the right. This is called a mono channel. And the recording of such sound is carried out only from one input device, for example, a conventional microphone.

However, when we need to add more different kinds of sounds/effects, this is where stereo sound comes to us. Its fundamental difference lies in the fact that the received sound does not go to both Left-Right channels, but specifically to the Left and separately to the Right [28]. Thus, the sound in the channels acquires a certain volume in the sound. For a more comparative analysis, you can see their display in Fig. 3.

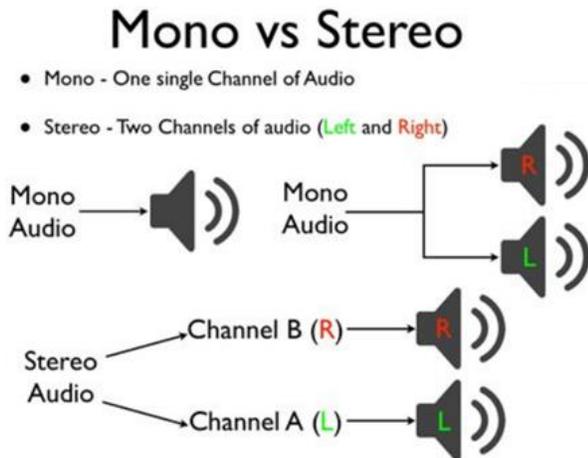


Fig. 3. Mono and stereo comparison.

And in order to see the difference for ourselves, we can take a different sound and compare their graphs, it is enough to display them through the Matplotlib library, which will allow us to do this [29]. To do this, we import and take another example with a mono channel for comparison, this will be a barking dog. Now let's display in Fig. 4:

In the graph, we can see that mono sound is displayed as one color, when the colors are displayed differently in stereo. Also, to check the channel, you can write a function that, using .shape, will show us a mono or stereo channel. This will especially help in the analysis of the second dataset.

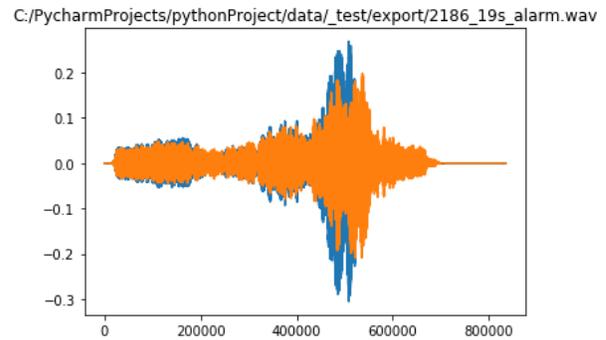
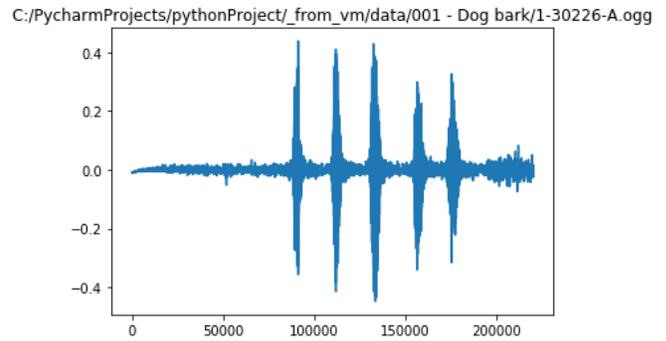


Fig. 4. Visualization of mono sound (upper), stereo sound (lower).

#### E. Proposed Model

There are deep learning techniques that can be applied in different areas as sound processing, images or video processing [30]. Urban environments, marked by their dynamic interactions and complexities, continuously emanate a diverse array of sounds, ranging from the benign murmurs of daily life to potentially dangerous noises that can indicate emergent situations or hazards [31]. Accurately discerning and classifying these dangerous sounds is not only paramount for the enhancement of urban safety but also imperative for proactive response mechanisms in smart cities. Traditional sound classification techniques often fall short in recognizing these transitory yet critical sounds due to their inherent limitations in capturing temporal relationships [32]. Enter the Bidirectional Long Short-Term Memory (BiLSTM) model, a sophisticated neural network architecture designed to navigate such challenges with unparalleled efficacy [33]. Fig. 5 demonstrates a flowchart of the proposed BiLSTM network for impulsive urban sound detection.

At its core, the Long Short-Term Memory (LSTM) is a form of Recurrent Neural Network (RNN) that addresses the vanishing gradient problem inherent in traditional RNNs. LSTMs are equipped with memory cells that can maintain information in memory for long periods, making them especially adept at tasks that require the understanding of long-term dependencies — a feature highly relevant to sound

sequences where past sounds can influence the characterization of present ones.

However, when dealing with dangerous urban sounds, which are often abrupt and embedded within larger, intricate auditory contexts, it becomes essential to understand the sound in relation to both its past and forthcoming sequences. This is where the bidirectional approach of the BiLSTM becomes invaluable. Instead of processing sequences in a unidirectional manner (from past to present), the BiLSTM simultaneously processes the data in both forward and backward directions. This bidirectional processing ensures that the model has access

to information from both before and after a particular time step, enabling a more comprehensive understanding of the sound's context.

In the context of dangerous urban sound classification, this means that a sudden loud crash, which could signify a vehicular accident or a structural collapse, is not just evaluated based on preceding sounds, but also by the sounds that follow it. Such a dual-context perspective can be crucial in distinguishing between, say, a harmless crash in a construction site versus a car collision that requires immediate attention.

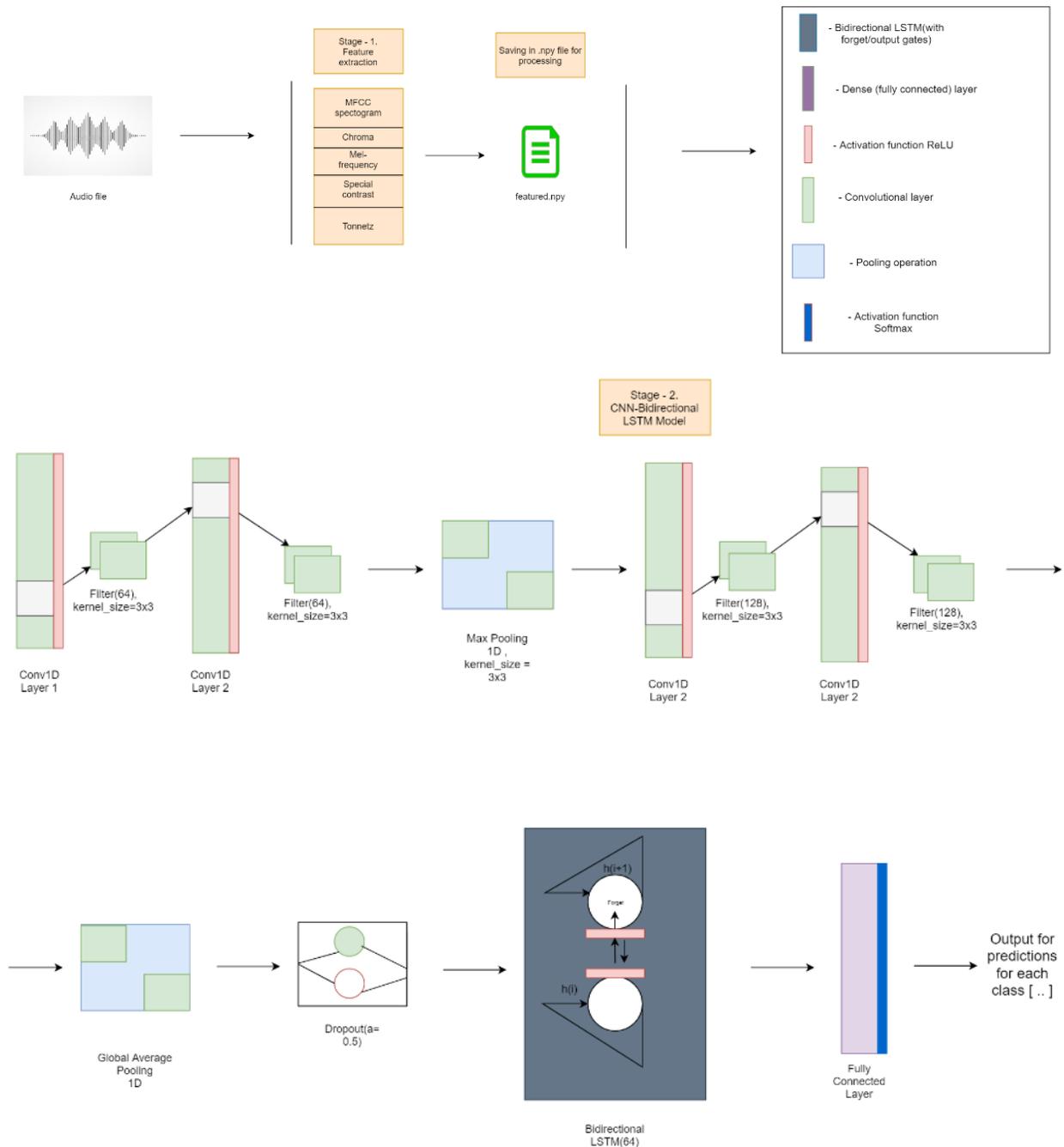


Fig. 5. Proposed Bidirectional LSTM network.

#### IV. EXPERIMENTAL RESULTS

Furthermore, the inherent structure of the BiLSTM, with its memory gates, allows for meticulous filtering of sound data, ensuring that only relevant information is retained for classification. This selective retention is especially crucial for urban environments where the soundscape is cluttered, and the distinction between dangerous and non-dangerous sounds can be razor-thin.

##### A. Results of the Proposed Model

The proposed Bidirectional Long Short-Term Memory model offers a groundbreaking approach to dangerous urban sound classification. Its ability to capture intricate temporal relationships from both past and future contexts, coupled with its adeptness at managing long-term dependencies, positions the BiLSTM as a frontrunner in the ongoing quest for creating safer and smarter urban ecosystems. As urban centers across the globe grapple with the challenges of increasing density and complexity, such advanced neural network architectures emerge not just as academic curiosities, but as essential tools for ensuring the well-being and safety of their inhabitants.

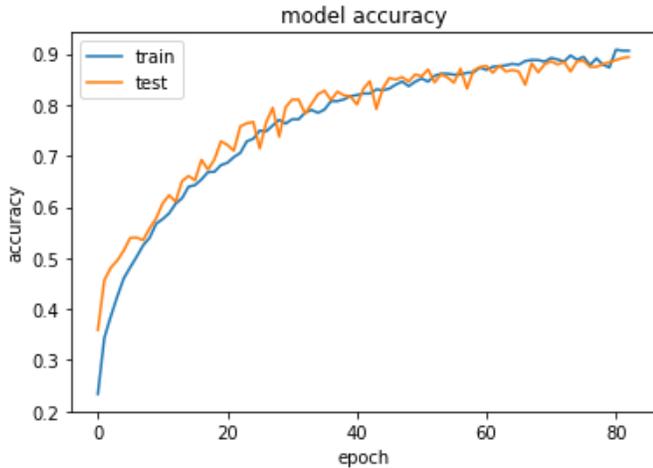


Fig. 6. Model training accuracy.

Further, this is a bidirectional LSTM network that immediately grows in parameters to as many as 352,330. This means that in theory there should be good results. The training time took about 871 seconds or about 14 minutes. Fig. 6 demonstrates the model training and test accuracy for 80 learning epochs.

Fig. 7 demonstrates model training loss in 80 learning epochs. The results show that, the proposed bidirectional LSTM network achieves to 90% accuracy, and 10% training loss, respectively.

Fig. 8 and Fig. 9 demonstrate test accuracy and test loss of the proposed model. Test results show that, the proposed model achieves 90% accuracy in model testing.

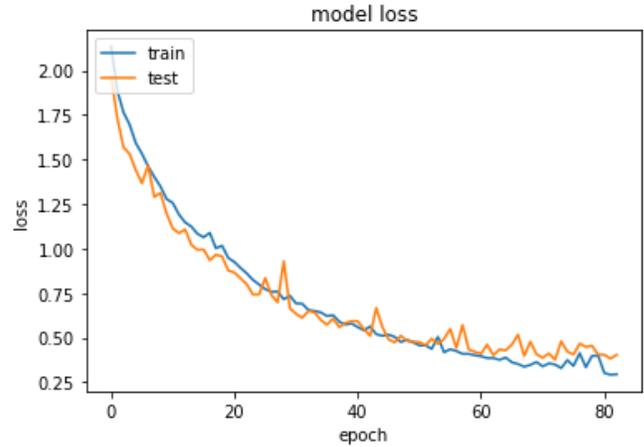


Fig. 7. Model training loss.

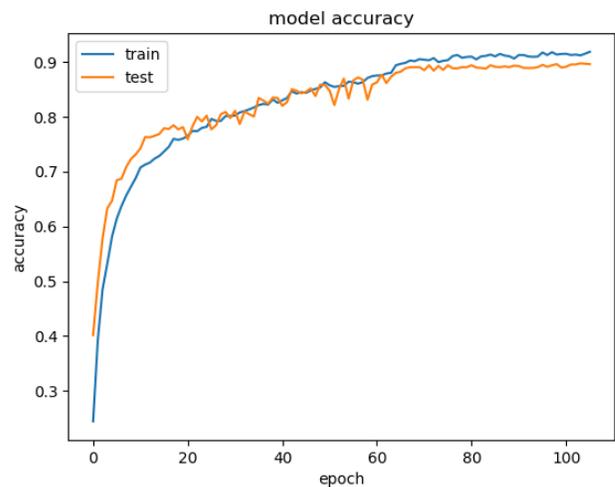


Fig. 8. Model test accuracy.

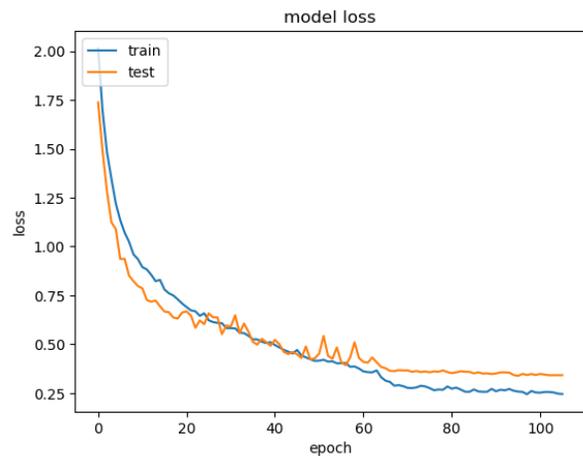


Fig. 9. Model test loss.

## V. DISCUSSION

The intricacies and dynamics of urban environments demand a profound, nuanced understanding, especially when delving into the auditory spectrum of these landscapes. Our exploration into the Bidirectional Long Short-Term Memory (BiLSTM) model for classifying dangerous urban sounds opens an array of discussions, both in the realm of neural network architectures and urban acoustics.

One of the most compelling findings from this research is the marked superiority of the BiLSTM model in classifying impulsive, dangerous sounds as compared to traditional sound classification techniques. The bidirectional nature of the model, which processes sound sequences both in forward and reverse temporal orders, demonstrates an inherent advantage in capturing the context of impulsive noises [34]. By concurrently assessing preceding and subsequent sounds, the model offers a panoramic view of the auditory environment, a perspective pivotal in discerning potential dangers in bustling urban soundscapes.

However, while the BiLSTM demonstrates significant promise, it's essential to address its limitations. Training a BiLSTM model, particularly with expansive urban sound datasets, can be computationally demanding [35]. The simultaneous processing of forward and backward sequences necessitates robust computational resources, which may not be readily available in all application scenarios, especially in real-time urban monitoring systems [36]. As cities move toward the vision of smart urbanism, the real-time processing of data becomes crucial. Future research should, therefore, look into optimizing the BiLSTM structure without compromising its classification prowess.

Another aspect worth reflecting upon is the diversity of the urban soundscape dataset employed in this study [37]. While the dataset was expansive, it was primarily curated from a limited number of urban settings. Urban soundscapes can significantly vary based on factors like cultural practices, architectural designs, traffic patterns, and even weather conditions. For the BiLSTM model to be universally applicable, it's imperative to train it with a more globally representative dataset, encompassing the myriad variations of urban environments. This would enhance the model's adaptability, ensuring its efficacy across diverse urban landscapes.

Furthermore, the human auditory system, despite its biological limitations, possesses a remarkable ability to discern sounds based on learned experiences and cultural contexts [38]. The sudden clang of pots in one culture might be dismissed as a benign household activity, while in another, it could be an alert for danger. Incorporating such cultural nuances and learned experiences into the BiLSTM model presents a challenge and an opportunity. The integration of these elements might enhance the model's sensitivity to context-specific dangerous sounds, making it even more aligned with human auditory perception.

Lastly, the ethical considerations of continuous urban sound monitoring need to be highlighted. While the primary intent is safety and rapid response to dangerous situations, the

omnipresent nature of sound monitoring systems can raise concerns related to privacy and surveillance. It becomes imperative for urban planners and policymakers to strike a balance, ensuring that the pursuit of safety doesn't infringe upon the privacy rights of city inhabitants.

In summation, this research underscores the transformative potential of the BiLSTM model in the realm of dangerous urban sound classification. The model's bidirectional processing, its adeptness at capturing temporal nuances, and its alignment with the holistic human perception of sounds make it an invaluable tool in the urban auditory toolkit. However, like all pioneering endeavors, this study raises as many questions as it seeks to answer. The computational demands of the model, the need for a more globally diverse dataset, the integration of cultural nuances, and the overarching ethical considerations form a rich tapestry of challenges and opportunities for future research.

As urban centers continue to burgeon and evolve, the imperative to understand, manage, and respond to their auditory landscapes becomes even more pronounced. The Bidirectional Long Short-Term Memory model, with its blend of technological sophistication and auditory acumen, emerges as a beacon in this journey, illuminating the path toward safer, smarter, and more responsive urban ecosystems. This research, albeit a single step, paves the way for a future where cities don't just listen but truly understand.

## VI. CONCLUSION

In an era where urban expanses are rapidly growing, manifesting themselves as the epicenters of human civilization, understanding the multifaceted dimensions of these environments is imperative. The auditory realm of cities, teeming with a symphony of sounds both benign and dangerous, necessitates an analytical lens equipped with both precision and depth. This research, centered on the Bidirectional Long Short-Term Memory (BiLSTM) model, underscores this very sentiment, offering a pioneering approach to the classification of dangerous urban sounds.

Our exploration into the BiLSTM model has illuminated its profound potential. By processing sound sequences in both forward and reverse temporal frames, the model imitates the holistic human perception of sounds, transcending the limitations of traditional classification techniques. This bidirectional prowess not only captures the intricate nuances of dangerous sounds but also provides a broader context, pivotal for accurate classification in bustling urban settings.

However, as is characteristic of any academic endeavor, this study also opens avenues for further exploration. While the BiLSTM model is undeniably potent, its computational demands, adaptability across diverse urban landscapes, and the integration of cultural and learned auditory nuances present challenges warranting future research. Moreover, the ethical dimensions of continuous urban sound monitoring, with potential implications for privacy and surveillance, underscore the need for a balanced approach, harmonizing safety with individual rights.

In conclusion, this research signifies a seminal step in the realm of urban sound classification. The BiLSTM model

emerges not merely as a technological marvel but as a testament to the convergence of neural network architectures and urban auditory science. As cities continue their inexorable march towards the future, tools like the BiLSTM will play a pivotal role, ensuring that these urban giants are not just expanses of concrete and steel, but responsive, adaptive, and safe ecosystems for all their inhabitants.

#### ACKNOWLEDGMENT

This work was supported by the research project — Development of a system for detecting and alerting dangerous events based on the audio analysis and machine learning funded by the Ministry of Science and Higher Education of the Republic of Kazakhstan. Grant No. IRN AP19175674.

#### REFERENCES

- [1] J. Bajzik, J. Prinosil, R. Jarina and J. Mekyska, “Independent channel residual convolutional network for gunshot detection,” *International Journal of Advanced Computer Science and Applications*, vol. 13, no.4, pp. 950-958, 2022.
- [2] K. M. Nahar, F. Al-Omari, N. Alhindawi and M. Banikhalf, “Sounds recognition in the battlefield using convolutional neural network,” *International Journal of Computing and Digital Systems*, vol. 11, no.1, pp. 189-198, 2022.
- [3] I. Estévez, F. Oliveira, P. Braga-Fernandes, M. Oliveira, L. Rebouta et al., “Urban objects classification using Mueller matrix polarimetry and machine learning,” *Optics Express*, vol. 30, no.16, pp. 28385-28400, 2022.
- [4] Z. Peng, S. Gao, Z. Li, B. Xiao, Y. Qian, “Vehicle safety improvement through deep learning and mobile sensing” *IEEE Network*, vol. 32, no.4, pp. 28-33, 2018.
- [5] Y. Wei, L. Jin, S. Wang, Y. Xu and T. Ding, “Hypoxia detection for confined-space workers: photoplethysmography and machine-learning techniques,” *SN Computer Science*, vol.3, no.4, pp.1-11, 2022.
- [6] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019). Applying face recognition in video surveillance security systems. In *Software Technology: Methods and Tools: 51st International Conference, TOOLS 2019, Innopolis, Russia, October 15–17, 2019, Proceedings 51* (pp. 271-280). Springer International Publishing.
- [7] K. Pawar, V. Attar, “Deep learning approaches for video-based anomalous activity detection,” *World Wide Web*, vol. 22, no.2, pp.571-601, 2019.
- [8] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. *Computers, Materials & Continua*, 74(3).
- [9] H. Zogan, I. Razzak, X. Wang, S. Jameel, G. Xu, “Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media,” *World Wide Web*, vol. 25, no.1, pp. 281-304, 2022.
- [10] C. Heipke, F. Rottensteiner, “Deep learning for geometric and semantic tasks in photogrammetry and remote sensing,” *Geo-spatial Information Science*, vol. 23, no.1, pp. 10-19, 2020.
- [11] Y. Arslan, H. Canbolat, “Sound based alarming based video surveillance system design,” *Multimedia Tools and Applications*, vol. 81, no.6, pp. 7969-7991, 2022.
- [12] A. Rajbanshi, D. Das, V. Udutalappally, R. Mahapatra, “DLeak: an IoT-based gas leak detection framework for smart factory,” *SN Computer Science*, vol. 3, no.4, pp. 1-12, 2022.
- [13] Y. Arslan and H. Canbolat, “Sound based alarming based video surveillance system design,” *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 7969-7991, 2022.
- [14] R. Sun, Q. Cheng, F. Xie, W. Zhang, T. Lin et al., “Combining machine learning and dynamic time wrapping for vehicle driving event detection using smartphones,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no.1, pp.194-207, 2019.
- [15] G. Chen, F. Wang, S. Qu, K. Chen, J. Yu et al., “Pseudo-image and sparse points: vehicle detection with 2D LiDAR revisited by deep learning-based methods,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no.12, pp. 7699-7711, 2020.
- [16] S. U. Amin, M. S. Hossain, G. Muhammad, M. Alhussein, M. A. Rahman, “Cognitive smart healthcare for pathology detection and monitoring,” *IEEE Access*, vol. 7, no.1, pp. 10745-10753, 2019.
- [17] Altayeva, A. B., Omarov, B. S., Aitmagambetov, A. Z., Kendzhaeva, B. B., & Burkitbayeva, M. A. (2014). Modeling and exploring base station characteristics of LTE mobile networks. *Life Science Journal*, 11(6), 227-233.
- [18] I. H. Peng, P. C. Lee, C. K. Tien, J. S. Tong, “Development of a cycling safety services system and its deep learning bicycle crash model,” *Journal of Communications and Networks*, vol. 24, no. 2, pp. 246-263, 2022.
- [19] L. Kou, “A review of research on detection and evaluation of the rail surface defects,” *Acta Polytechnica Hungarica*, vol. 19, no.3, pp. 167-186, 2022.
- [20] L. M. Bine, A. Boukerche, L. B. Ruiz, A. A. Loureiro, “Leveraging urban computing with the internet of drones,” *IEEE Internet of Things Magazine*, vol. 5, no.1, pp. 160-165, 2022.
- [21] S. Khan, L. Alarabi and S. Basalamah, “Toward smart lockdown: a novel approach for COVID-19 hotspots prediction using a deep hybrid neural network,” *Computers*, vol. 9, no. 4, pp. 1-16, 2020.
- [22] M. Dua, D. Makhija, P. Manasa and P. Mishra, “A CNN–RNN–LSTM based amalgamation for Alzheimer’s disease detection,” *Journal of Medical and Biological Engineering*, vol. 40, no. 5, pp. 688-706, 2020.
- [23] H. Gill, O. Khalaf, Y. Alotaibi, S. Alghamdi and F. Alassery, “Multi-model CNN-RNN-LSTM based fruit recognition and classification,” *Intelligent Automation & Soft Computing*, vol. 33, no. 1, pp. 637-650, 2022.
- [24] K. Chandriah and R. Naraganahalli, “RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting,” *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26145-26159, 2021.
- [25] S. Hansun and J. Young, “Predicting LQ45 financial sector indices using RNN-LSTM,” *Journal of Big Data*, vol. 8, no. 1, pp. 1-13, 2021.
- [26] Y. Xue, P. Shi, F. Jia, H. Huang, “3D reconstruction and automatic leakage defect quantification of metro tunnel based on SfM-Deep learning method,” *Underground Space*, vol. 7, no.3, pp. 311-323, 2022.
- [27] L. Zhang, L. Yan, Y. Fang, X. Fang, X. Huang, “A machine learning-based defensive alerting system against reckless driving in vehicular networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no.12, pp.12227-12238, 2019.
- [28] A. M. Youssef, B. Pradhan, A. Dikshit, M. M. Al-Katheri, S. S. Matar et al., “Landslide susceptibility mapping using CNN-1D and 2D deep learning algorithms: comparison of their performance at Asir Region, KSA,” *Bulletin of Engineering Geology and the Environment*, vol. 81, no.4, pp. 1-22, 2022.
- [29] S. Asadianfam, M. Shamsi, A. Rasouli Kenari, “Hadoop Deep Neural Network for offending drivers,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no.1, pp. 659-671, 2022.
- [30] Tursynova, A., & Omarov, B. (2021, November). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. In *2021 16th International Conference on Electronics Computer and Computation (ICECCO)* (pp. 1-4). IEEE.
- [31] D. K. Dewangan, S. P. Sahu, “Deep learning-based speed bump detection model for intelligent vehicle system using raspberry Pi,” *IEEE Sensors Journal*, vol. 21, no.3, pp. 3570-3578, 2020.
- [32] Z. Fang, B. Yin, Z. Du and X. Huang, “Fast environmental sound classification based on resource adaptive convolutional neural network,” *Scientific Reports*, vol. 12, no. 1, pp. 1-18, 2022.
- [33] V. Gughani, R. K. Singh, “Analysis of deep learning approaches for air pollution prediction,” *Multimedia Tools and Applications*, vol. 81, no.4, pp. 6031-6049, 2022.

- [34] Omarov, B., Altayeva, A., Turganbayeva, A., Abdulkarimova, G., Gusmanova, F., Sarbasova, A., ... & Omarov, N. (2019). Agent based modeling of smart grids in smart cities. In *Electronic Governance and Open Society: Challenges in Eurasia: 5th International Conference, EGOSE 2018, St. Petersburg, Russia, November 14-16, 2018, Revised Selected Papers 5* (pp. 3-13). Springer International Publishing.
- [35] H. Kyle, P. Agarwal, J. Zhuang, "Monitoring misinformation on Twitter during crisis events: a machine learning approach," *Risk Analysis*, vol. 42, no.8, pp. 1728-1748, 2022.
- [36] M. Esmail Karar, O. Reyad, A. Abdel-Aty, S. Owyed, M. F. Hassan, "Intelligent iot-aided early sound detection of red palm weevils," *Computers, Materials & Continua*, vol. 69, no.3, pp. 4095-4111, 2021.
- [37] T. Thomas Leonid and R. Jayaparvathy, "Classification of elephant sounds using parallel convolutional neural network," *Intelligent Automation & Soft Computing*, vol. 32, no.3, pp. 1415-1426, 2022.
- [38] Z. Ma, G. Mei, , F. Piccialli, "Machine learning for landslides prevention: a survey," *Neural Computing and Applications*, vol. 33, no.17, pp. 10881-10907, 2021.