

Emotion-Based Voice Control for IoT: Enhancing Smart Device Interaction with Speech Emotion Classification*

Ihor Shevchuk^{1†}, Iryna Dumyn^{1†}

¹ Lviv Polytechnic National University, 12 Stepana Bandery Str., Lviv, 79000, Ukraine

Abstract

Speech emotion recognition (SER) is essential for enhancing human-computer interaction, especially within voice-controlled IoT systems. This study explores various machine learning and deep learning approaches for classifying emotions from speech signals, utilizing features such as Mel-Frequency Cepstral Coefficients (MFCCs). The research evaluates the performance of traditional models, including classical machine learning algorithms such as Naïve Bayes, Logistic Regression, Decision Trees, and Random Forest, alongside deep learning models - Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM). A five-fold cross-validation strategy is employed to ensure robust performance assessment. The experimental results demonstrate that CNN-based models achieve the highest accuracy, followed closely by LSTM networks, highlighting the effectiveness of deep learning in capturing temporal and spectral patterns in speech. Traditional machine learning models also show competitive performance, emphasizing the importance of feature extraction techniques. The study discusses the challenges of real-time deployment, the impact of dataset size, and the need for robust models that can generalize across speakers and environments. Future work will focus on optimizing deep learning architectures, integrating multimodal inputs, and improving model efficiency for real-time IoT applications. These advancements will contribute to the development of more intelligent and responsive voice-controlled systems capable of recognizing and adapting to human emotions.

Keywords

IoT, machine learning, speech recognition, NLU, classification algorithms, neural networks, TensorFlow.

1. Background

The rapid expansion of the Internet of Things (IoT) has transformed the way people interact with smart technologies. Traditionally, voice control has played a crucial role in improving accessibility, allowing users to operate IoT systems using simple voice commands. However, conventional voice-controlled systems lack emotional intelligence [1], meaning they respond to commands without understanding the speaker's mood or emotional state. This limitation results in robotic and non-personalized interactions, which can reduce user satisfaction.

Emotion classification from speech offers a novel way to enhance voice-controlled IoT systems by making them more adaptive and responsive. Humans naturally adjust their communication style based on emotions, and enabling IoT devices to do the same can significantly improve user experience. Advanced machine learning (ML) and deep learning (DL) algorithms can analyze voice signals and classify emotions such as happiness, sadness, anger, and neutrality [2]. Integrating such a system into IoT devices enables them to react differently based on the detected emotions rather than following rigid command structures.

For example, a smart home assistant could detect frustration in a user's voice and provide a calming response or suggest relaxation music. Similarly, a healthcare IoT device can monitor the emotional state of elderly users and alert caregivers if signs of stress or depression are detected. In professional environments, emotion-aware IoT systems can enhance customer service by detecting dissatisfaction in customer voices and adjusting the system's behavior accordingly.

The Second International Conference of Young Scientists on Artificial Intelligence for Sustainable Development (YAISD), May 8-9, 2025, Ternopil, Ukraine

* Corresponding author.

† These authors contributed equally.

✉ ihor.o.shevchuk@lpnu.ua (I. Shevchuk); iryna.b.shvorob@lpnu.ua (I. Dumyn);

ORCID [0000-0002-8319-8934](https://orcid.org/0000-0002-8319-8934) (I. Shevchuk); [0000-0001-5569-2647](https://orcid.org/0000-0001-5569-2647) (I. Dumyn)

Developing an accurate emotion classification model[3] requires analyzing various acoustic features such as pitch, energy, Mel-Frequency Cepstral Coefficients (MFCCs)[4-5], and spectral characteristics. These features help machine learning models differentiate between emotions. The implementation of Convolutional Neural Networks (CNNs)[6-7] and Long Short-Term Memory (LSTM)[8-9] networks has significantly improved speech emotion classification, achieving high accuracy in real-world applications. With growth in this direction new packages for processing voice signals are becoming very popular(i.e. librosa[10], speech_recognition[11], etc.)

However, challenges remain in creating real-time emotion detection systems that are robust to different accents, background noises, and variations in speech intensity. Additionally, processing speech data on IoT devices presents hardware constraints, as these systems often have limited processing power compared to cloud-based solutions.

The “Global Voice and Speech Recognition Market”[12] was valued at USD 20.25 billion in 2023 and is expected to grow at a CAGR of 14.6% from 2024 to 2030 as displayed on figure 1. This growth is driven by technological advancements and the increasing adoption of advanced electronic devices. Voice-activated biometrics enhance security by granting access only to authenticated users for transactions, contributing significantly to market expansion. The rising demand for voice-driven navigation systems and workstations is fueling growth in both hardware and software segments. Additionally, the integration of voice-enabled in-car infotainment systems is gaining traction worldwide, driven by the implementation of “hands-free” regulations in various countries that restrict mobile phone use while driving.

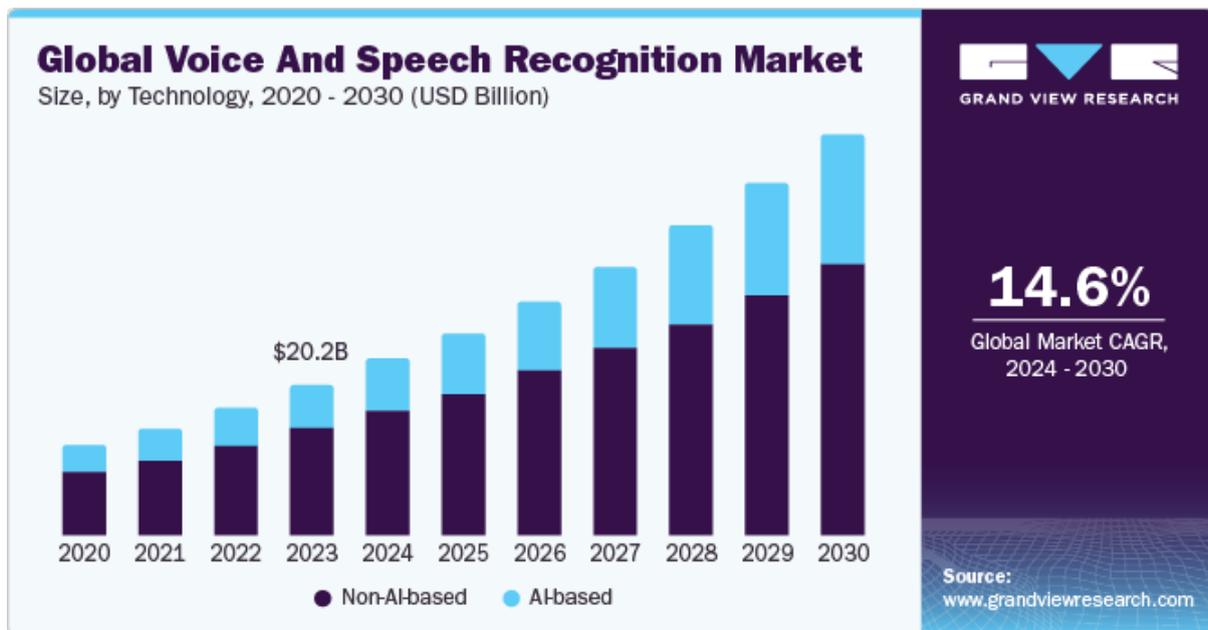


Figure 1: Growth of Global Voice and Speech Recognition Market[12]

This article explores how speech emotion classification can be integrated into IoT-based voice control systems. We discuss machine learning models, feature extraction techniques, and dataset requirements, along with challenges in real-world implementation. By the end of this study, we aim to build an effective approach for emotion classification that can be used in modern IoT systems leading to a more natural and personalized interaction experience.

In lecture notes Review of Automatic Speech Recognition Systems for Ukrainian and English Language [13] authors Dumyn A, Fedushko S., Syerov Y. explored various classification approaches for speech-based emotion recognition, including classical machine learning methods such as Logistic Regression and Naïve Bayes, as well as deep learning models like CNNs and LSTMs. The study investigates the effectiveness of Mel-Frequency Cepstral Coefficients (MFCCs) as features for emotion classification and evaluates the performance of different models using accuracy and other relevant metrics. By analyzing these methods, the research provides insights into how speech emotion recognition can enhance human-machine interaction, complementing advancements in automatic

speech recognition systems. This study shows high efficiency of such methods and suggests them to be used for further investigation.

The study [14] strongly advocates for the use of recurrent neural networks (RNNs) in processing human speech recordings, emphasizing their ability to capture specific features in voice data. It highlights RNN-based models as particularly effective for analyzing voice messages, as they can retain contextual information and improve the accuracy of speech emotion recognition systems.

2. Task Statement

The goal of this research is to develop an emotion classification system that can analyze human speech, detect emotions, and integrate the results into voice-controlled IoT devices. Current voice control technologies primarily rely on recognizing specific words and commands, but do not account for emotional context. This project aims to bridge that gap by building an emotion-aware IoT system that dynamically adjusts its responses based on the user's emotional state.

Main Research Objectives:

1. Dataset Collection and Preprocessing – Gather a high-quality emotional speech dataset such as RAVDESS[15], CREMA-D[16], TESS[17], etc.
2. Feature Extraction – Extract key audio features like MFCCs, pitch, chroma features, and spectral centroid to train emotion classification models.
3. Model Development – Design and train a deep learning model using classical ML algorithms, CNNs, LSTMs.
4. Performance Evaluation – Measure the model's accuracy, robustness to noise, and real-time inference capability in different environments.

By achieving these objectives, we can create an IoT voice control system that adapts to users' emotions, improving personalization and user satisfaction. The system can be applied in smart homes, healthcare monitoring, customer service, and assistive technology for individuals with disabilities.

Emotion classification from speech is a rapidly growing field with applications in human-computer interaction, healthcare, call center analytics, and smart IoT devices. Understanding emotions from audio signals allows systems to interpret user intent, respond appropriately, and create more natural and engaging interactions. In this research, we focus on automatically classifying emotions from voice recordings using various machine learning and deep learning models.

Traditional methods for speech-based emotion recognition typically depend on handcrafted features derived from raw audio signals. These features often include pitch, energy, Mel-Frequency Cepstral Coefficients (MFCCs), and various spectral characteristics.

In contrast, modern deep learning techniques such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks can automatically identify meaningful patterns from speech data. Despite the advancements in deep learning, classical machine learning algorithms like Logistic Regression and Naïve Bayes continue to play an important role due to their simplicity, interpretability, and suitability for smaller datasets.

This study evaluates a range of classification techniques, including:

- Logistic Regression – A basic linear classifier that models the relationship between extracted audio features and corresponding emotional states.
- Naïve Bayes Classifier – A fast, probabilistic model based on Bayes' theorem, well-suited for tasks such as text and speech classification.
- Convolutional Neural Networks (CNNs) – Deep learning models that extract spatial features from spectrogram representations of speech.
- Long Short-Term Memory (LSTM) Networks – A form of Recurrent Neural Network (RNN) capable of capturing temporal dependencies in sequential speech data.

Through comparative analysis, this study aims to identify the most effective approach for real-time emotion classification. Each model offers distinct advantages: Logistic Regression and Naïve Bayes provide fast and lightweight solutions, while CNNs and LSTMs deliver greater accuracy by capturing complex patterns, albeit with higher computational demands.

This work lays the foundation for building real-time emotion-aware IoT devices. Speech-controlled systems that understand human emotions can enhance user experiences in smart homes, virtual assistants, and automated customer service.

After a command is recorded by the user, a message is sent to the local control unit/cloud(AWS IoT). Next step is to understand speech and classify emotions. An AI agent creates commands for IoT devices with some adjustments to desired based on those inputs(according to classified emotion) as shown on figure 2.

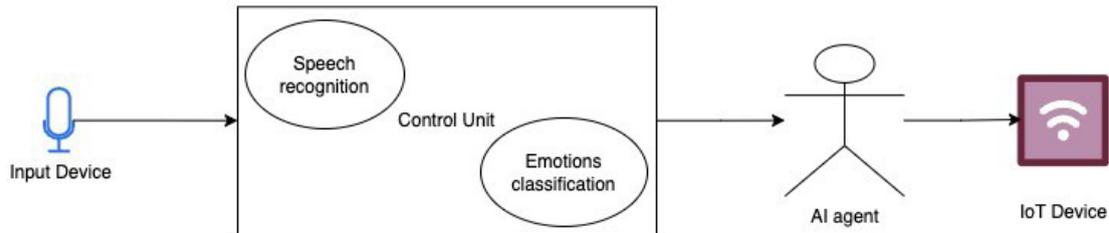


Figure 2: Diagram of real-time emotion-aware IoT system.

3. Method and results

3.1 Dataset collection

For this article Tess[17] dataset was chosen. Toronto emotional speech set data is an open source dataset that contains a set of target words that were spoken in a specific manner “Say the word _” by 2 actresses aged 26 and 64 years. Recordings were grouped by 7 emotions - anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral(table 1). All recordings are stored in WAV format.

Table 1

Data classes distribution

Emotion	Number of samples
Anger	400
Disgust	400
Fear	400
Happiness	400
Pleasant surprise	400
Sadness	400
Neutral	400

3.2 Feature extraction

To process the dataset MFCC algorithm was used. Mel-Frequency Cepstral Coefficients (MFCCs) are one of the most widely used features for speech and audio processing. They are designed to mimic the way the human auditory system perceives sound, making them highly effective for speech recognition, speaker identification, and emotion classification.

Raw audio waveforms contain a lot of information, but not all of it is useful for classification. MFCC helps by extracting key features that represent how humans hear and process sounds. Human Perception-Based – MFCCs use a Mel scale, which reflects how humans perceive pitch. Compact Representation – Instead of using the full audio waveform, MFCCs extract a smaller set of meaningful numbers.

MFCC extraction process can be represented as a single general formula that captures the transformation from the raw speech signal to the final MFCC features.

$$C_n = \sum_{m=1}^M \log \left(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k) \cos \left[\frac{\pi m(2m+1)}{2M} \right] \right) \quad (1)$$

where

$X_m(k)$ is the Fourier Transform of the framed speech signal.

$H_m(k)$ represents the Mel filterbank applied in the frequency domain.

The inner sum computes the energy output of each Mel filter.

The logarithm models human loudness perception.

The outer sum applies the Discrete Cosine Transform (DCT)[18] to decorrelate the features and generate MFCCs.

C_n are the final MFCC coefficients, used for speech and emotion recognition.

MFCC are effective for Speech Analysis – They emphasize frequency ranges important for understanding speech and emotions. For our experiment was used sample rate of 40Hz resulting in dataset with shape 5600, 40.

3.3 Model Development

For our experiment few classical ML algorithms were chosen as well as one architecture of CNN and LSTM for comparison.

- **Logistic Regression** is a straightforward classification technique that predicts class probabilities by transforming a linear combination of input features using a sigmoid function. It is known for its interpretability and effectiveness in handling data that is linearly separable, though its performance may drop with complex, non-linear patterns.

- **Naïve Bayes** relies on probabilistic reasoning and assumes independence among features. It is highly efficient, requires little training data, and is suitable for structured datasets. However, its simplifying assumptions may reduce accuracy when there is strong correlation between features.

- **Decision Trees** operate by iteratively partitioning the data based on feature importance, forming a tree-like model structure. They are easy to understand and handle non-linear data well but are often prone to overfitting, especially when the dataset is noisy or small.

- **Random Forests** build on Decision Trees by generating an ensemble of them and aggregating their outputs. This approach boosts predictive accuracy and reduces overfitting risks, though at the cost of increased computational load.

- **Convolutional Neural Networks (CNNs)** are deep learning models well-suited for analyzing spatial patterns in data such as spectrograms. They extract complex features through layers of convolutions, but their effectiveness often depends on the availability of large datasets and substantial computational resources.

- **Long Short-Term Memory (LSTM)** networks, a variant of Recurrent Neural Networks (RNNs), are designed to capture long-range dependencies in sequential data. These models are especially useful in applications like speech recognition, although their training process can be slow and resource-intensive due to their layered structure.

3.4 Performance Evaluation

In this research, a 5-fold cross-validation[19] technique was employed to assess the performance of various models used for speech emotion classification. The dataset was divided into five subsets of equal size, with each model trained on four of these subsets and evaluated on the remaining one. This cycle was repeated five times so that each subset served as the test set once, ensuring balanced and unbiased evaluation.

This cross-validation approach enhances the reliability of the results by minimizing the impact of data partitioning. By averaging the results across all five folds, we obtain a more consistent estimate of the model's general performance. Metrics such as accuracy, precision, recall, and F1-score were used to quantify model effectiveness across folds.

Such a strategy is especially beneficial for smaller datasets, as it allows maximum utilization of available data for training while still enabling comprehensive performance evaluation. Compared to a basic train-test split, this method offers more robust insights into how well the model is likely to perform on unseen data.

Results of classification are present in table 1.

Table 2
Classification results for emotions recognition

Model Name	K-Fold 1	K-Fold 2	K-Fold 3	K-Fold 4	K-Fold 5	AVG
Logistic regression	0.9875	0.9938	0.9844	0.9979	0.9844	0.9896
Naive Bayes	0.9866	0.9902	0.9884	0.9911	0.9929	0.9898
Decision Tree	0.9792	0.9896	0.9813	0.9792	0.9751	0.9809
Random Forest	0.9958	1	0.9979	1	0.9938	0.9975
CNN	0.9982	0.9991	1	0.9982	1	0.9991
LSTM	0.9866	0.9902	0.9884	0.9911	0.9929	0.9898

The table includes accuracy values for each fold and the average accuracy across all five folds, providing insight into the consistency and effectiveness of each approach.

Despite promising results in emotion recognition, integrating these models into real-world IoT environments remains a challenge. Many IoT devices operate with limited computational resources, making it difficult to deploy large deep learning models without optimization. Real-time emotion recognition further increases the complexity, as it demands both low-latency processing and high accuracy. Issues such as memory constraints, power consumption, and network connectivity must be considered when designing models for edge deployment. Techniques like model pruning, quantization, and edge-cloud collaboration are potential solutions to address these hardware limitations. Analysis of emotion control is given in the works [22-25].

The results indicate that deep learning models, particularly CNN and LSTM, achieved the highest classification accuracy. The CNN model demonstrated near-perfect performance with an average accuracy of 0.9991, suggesting its strong ability to extract meaningful patterns from speech features. Similarly, LSTM, designed for sequential data, performed competitively with an average accuracy of 0.9888.

Among traditional machine learning models, Logistic Regression and Naïve Bayes showed comparable performance, with average accuracies of 0.9896 and 0.9889, respectively. These results suggest that even simple classifiers can achieve high accuracy when trained on well-processed speech features. The Decision Tree model had the lowest accuracy at 0.9809, highlighting its tendency to overfit and perform inconsistently across different folds. Random Forest, an ensemble of decision trees, improved upon this with an average accuracy of 0.9975, demonstrating its ability to generalize better.

Overall, CNN and LSTM outperformed other methods, confirming that deep learning models are highly effective for emotion recognition tasks. However, traditional models like Logistic Regression and Naïve Bayes still achieved competitive results, making them viable options for real-time applications where computational efficiency is a priority.

4. Conclusions and Further Work

4.1 Conclusions

The results of the speech emotion recognition experiments demonstrate that deep learning models, particularly CNN and LSTM, achieve the highest classification performance, with the CNN model reaching an average accuracy of 0.9991 and the LSTM model following closely at 0.9888. Traditional machine learning models, such as Logistic Regression (0.9896) and Naïve Bayes (0.9889), also performed well, proving that well-engineered feature extraction can enable simpler models to compete with more complex architectures. The Decision Tree model had the lowest accuracy (0.9809),

highlighting its limitations in capturing the intricate patterns within speech data, whereas the Random Forest model showed significant improvement (0.9975) due to its ensemble nature.

The results suggest that while deep learning provides superior performance, traditional machine learning models can still be highly effective for speech emotion recognition when feature extraction is carefully designed. Furthermore, the small performance gap between traditional and deep learning models indicates that high-quality feature engineering can compensate for the lack of end-to-end learning.

4.2 Further Work

Although the models achieved high accuracy, several areas require further research and development. Expanding the dataset with more diverse emotional expressions and speakers would improve model generalization and robustness. Another promising direction is optimizing deep learning architectures by experimenting with hybrid approaches, such as combining CNNs and LSTMs or incorporating attention mechanisms to enhance sequence modeling.

A major challenge remains the real-time implementation of deep learning models in IoT applications, as they require significant computational resources. The authors [20] analyse methods of processing IoT data that can be useful for reducing those expenses. Future research should explore model compression, quantization, and lightweight architectures to make real-time deployment feasible. This research [21] could benefit from exploring new types of databases for storing such data. The authors of the study discuss Graph Databases as a novel approach for handling this kind of information. Additionally, transfer learning from pre-trained speech models could improve classification accuracy while reducing the need for large labeled datasets.

Moreover, environmental noise and speaker variability pose significant challenges in real-world applications. Future efforts should focus on developing noise-robust models using domain adaptation techniques and data augmentation methods. Exploring multimodal approaches that integrate facial expressions, physiological signals, or contextual information could further enhance the accuracy of emotion recognition systems.

By addressing these areas, emotion classification models can be refined for deployment in human-computer interaction, virtual assistants, and intelligent IoT systems, making voice-based emotion recognition a more practical and reliable tool in various real-world applications.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using those services, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

1. References

- [1] Ezzameli, K., & Mahersia, H. (2023). Emotion recognition from unimodal to multimodal analysis: A review. *Information Fusion*, 101, 101847. <https://doi.org/10.1016/j.inffus.2023.101847>
- [2] Norval, M., & Wang, Z. (2024). Speech Emotion Recognition using Hybrid Architectures. *International Journal of Computing*, 23(1), 1-10. <https://doi.org/10.47839/ijc.23.1.3430>
- [3] Jain, S., Basu, S., Ray, A., & Das, R. (2023). Impact of irritation and negative emotions on the performance of voice assistants: Netting dissatisfied customers' perspectives. *International Journal of Information Management*, 68, 102662. <https://doi.org/10.1016/j.ijinfomgt.2023.102662>
- [4] Abdul, Z. Kh., & Al-Talabani, A. K. (2022). Mel frequency cepstral coefficient and its applications: A review. *IEEE Access*, 10, 122136 - 122158. <https://doi.org/10.1109/ACCESS.2022.3223444>
- [5] Dwivedi, D., Ganguly, A., & Haragopal, V. V. (2023). Contrast between simple and complex classification algorithms. In *Advances in Intelligent Systems and Computing* (Vol. 155, pp. 123-135). Elsevier. <https://doi.org/10.1016/B978-0-323-91776-6.00016-6>

- [6] IBM. (n.d.). Convolutional Neural Networks. IBM Think. Retrieved from <https://www.ibm.com/think/topics/convolutional-neural-networks>.
- [7] Sung, W.-T., Kang, H.-W., & Hsiao, S.-J. (2023). Speech recognition via CTC-CNN model. *Computers, Materials & Continua*, 70(2), 1941-1954. <https://doi.org/10.32604/cmc.2023.040024>
- [8] Senthilkumar, N., Karpakam, S., Gayathri Devi, M., Balakumaresan, R., & Dhilipkumar, P. (2021). Speech emotion recognition based on Bi-directional LSTM architecture and deep belief networks. *Materials Today: Proceedings*, 43, 2135-2140. <https://doi.org/10.1016/j.matpr.2021.12.246>
- [9] Francis, N., Suhaimi, H., & Abas, E. (2023). Classification of Sprain and Non-sprain Motion using Deep Learning Neural Networks for Ankle Sprain Prevention. *International Journal of Computing*, 22(2), 159-169. <https://doi.org/10.47839/ijc.22.2.3085>
- [10] Quinn, C. A., Burns, P., Gill, G., Baligar, S., Snyder, R. L., Salas, L., Goetz, S. J., & Clark, M. L. (2022). Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data. *Ecological Indicators*, 138, 108831. <https://doi.org/10.1016/j.ecolind.2022.108831>
- [11] Patel, A., & Sharma, D. (2023). Automatic speech emotion recognition using deep learning. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2023.116030>
- [12] Grand View Research. (2023). Voice Recognition Market Size, Share & Trends Analysis Report. Retrieved from <https://www.grandviewresearch.com/industry-analysis/voice-recognition-market>.
- [13] Dumyn, A., Fedushko, S., Syerov, Y. (2024). Review of Automatic Speech Recognition Systems for Ukrainian and English Language. In: Štarchoň, P., Fedushko, S., Gubíniová, K. (eds) Data-Centric Business and Applications. Lecture Notes on Data Engineering and Communications Technologies, vol 212. Springer, Cham. https://doi.org/10.1007/978-3-031-60815-5_15
- [14] Basystiuk, O., Shakhovska, N., Bilynska, V., Syvokon, O., Shamuratov, O., & Kuchkovskiy, V. (2021). The Developing of the System for Automatic Audio to Text Conversion. In IT&AS (pp. 1-8).
- [15] Issa, D., Demirci, M. F., & Yazici, A. (2020). Speech emotion recognition with deep convolutional neural networks. *Biomedical Signal Processing and Control*, 63, 101894. <https://doi.org/10.1016/j.bspc.2020.101894>
- [16] Kanna, P. R., & Kumararaja, V. (2024). Enhancing speech emotion detection with Windowed Long-Term Average Spectrum and Logistic-Rectified Linear Unit. *Engineering Applications of Artificial Intelligence*, 113, 109103. <https://doi.org/10.1016/j.engappai.2024.109103>
- [17] Lok, E.J. (2017). Toronto Emotional Speech Set (TESS). Kaggle. Retrieved from <https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess>.
- [18] Wanhammar, L. (2001). Digital Signal Processing. In *Signal Processing and Linear Systems* (pp. 1-30). Elsevier. <https://doi.org/10.1016/B978-0-12734530-7/50003-9>
- [19] Spangler, W. D., Gupta, A., Kim, D. H., & Nazarian, S. (2013). Developing and validating historiometric measures of leader individual differences by computerized content analysis of documents. *The Leadership Quarterly*, 24(1), 5-18. <https://doi.org/10.1016/j.leaqua.2012.11.002>
- [20] O. Duda, V. Kochan, N. Kunanets, O. Matsiuk, V. Pasichnyk, A. Sachenko, T. Pytlenko, "Data Processing in IoT for Smart City Systems," The 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2019), 18-21 September, 2019, Metz, France, vol. 1, pp. 96-99.
- [21] Dumyn I., Basystiuk O., Dumyn A. Graph-Based Approaches for Multimodal Medical Data Processing // CEUR Workshop Proceedings. – 2024. – Vol. 3892: Proceedings of the 7th International Conference on Informatics & Data-Driven Medicine IDDM 2024, Birmingham, United Kingdom, November 14-16, 2024. – P. 337-348.
- [22] R. Gramyak, H. Lipyanina-Goncharenko, A. Sachenko, T. Lendyuk, D. Zahorodnia. Intelligent Method of a Competitive Product Choosing based on the Emotional Feedbacks Coloring. March 24–26, 2021, pp. 346-357. <http://ceur-ws.org/Vol-2853/paper31.pdf>.
- [23] Lipianina-Honcharenko, K., Savchyshyn, R., Sachenko, A., Chaban, A., Kit, I., & Lendiuk, T. (2022). Concept of the Intelligent Guide with AR Support. *International Journal of Computing*, 21(2), 271-277. <https://doi.org/10.47839/ijc.21.2.2596>
- [24] Francis, N., Suhaimi, H., & Abas, E. (2023). Classification of Sprain and Non-sprain Motion using Deep Learning Neural Networks for Ankle Sprain Prevention. *International Journal of Computing*, 22(2), 159-169. <https://doi.org/10.47839/ijc.22.2.3085>
- [25] Norval, M., & Wang, Z. (2024). Speech Emotion Recognition using Hybrid Architectures. *International Journal of Computing*, 23(1), 1-10. <https://doi.org/10.47839/ijc.23.1.3430>