JOURNAL OF MODERN SCIENCE

SPECIAL ISSUE

3/57/2024

www.jomswsge.com

DOI: doi.org/10.13166/jms/191144

EWA GUZ
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0000-0002-0507-2172

KONRAD NIDERLA
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0000-0003-1280-0622

GRZEGORZ KATA
WSEI University in Lublin, Poland
ORCID iD: orcid.org/0000-0001-8762-9378

ADVANCING AUTISM THERAPY: EMOTION ANALYSIS USING REHABILITATION ROBOTS AND AI FOR CHILDREN WITH ASD

POSTĘP W TERAPII AUTYZMU: ANALIZA EMOCJI PRZY UŻYCIU ROBOTÓW REHABILITACYJNYCH I SZTUCZNEJ INTELIGENCJI DLA DZIECI Z ASD

Abstract

Emotion analysis is a key component in understanding the unique communication patterns and emotional states of children with autism spectrum disorder (ASD). These children often struggle with traditional forms of expressing emotions, which presents a challenge for themselves and their therapists. Facial expression analysis techniques, supported by modern technologies such as machine learning and artificial intelligence, enable more accurate identification of subtle signals that may go unnoticed by human observers. The introduction of rehabilitation robots and emotion analysis software based on the analysis of facial expressions and gestures opens up new possibilities for individualizing therapy, adapting it to the child's specific reactions and needs. In this way, the use of these tools not only increases the effectiveness of treatment but also helps build more trusting therapeutic relationships, which is the basis for adequate support for the development of children with ASD. Regular monitoring of progress and modifying therapeutic approaches, supported by automation and data analysis, is essential to more effective and empathetic care for children with developmental disorders. However, the journey does not end here. Further research is necessary to develop and improve emotion analysis techniques for use in rehabilitation robots and their impact on the effectiveness of therapy for young patients.

Streszczenie

Analiza emocji odgrywa kluczową rolę w zrozumieniu unikalnych wzorców komunikacji i stanów emocjonalnych dzieci z zaburzeniami ze spektrum autyzmu (ASD). Dzieci te często napotykają trudności w tradycyjnych formach wyrażania emocji, co stanowi wyzwanie zarówno dla nich samych, jak i dla ich terapeutów. Techniki analizy wyrazu twarzy, wspomagane nowoczesnymi technologiami takimi jak uczenie maszynowe i sztuczna inteligencja, umożliwiają dokładniejszą identyfikację subtelnych sygnałów, które mogą pozostać niezauważone przez ludzkich obserwatorów. Wprowadzenie robotów rehabilitacyjnych i oprogramowania do analizy emocji opartej analizie wyrazu twarzy i gestów otwiera nowe możliwości indywidualizacji terapii, dostosowując ją do specyficznych reakcji i potrzeb dziecka. W ten sposób stosowanie tych narzędzi nie tylko zwiększa skuteczność terapii, ale także pomaga w budowaniu bardziej ufnych relacji terapeutycznych, co stanowi podstawę efektywnego wsparcia rozwoju dzieci z ASD. Regularny monitoring postępów i modyfikacje podejść terapeutycznych, wspierane automatyzacją i analizą danych, stanowią ważny krok w kierunku bardziej efektywnej i empatycznej opieki nad dziećmi z zaburzeniami rozwojowymi. Konieczne są dalsze badania w celu opracowania i udoskonalenia technik analizy emocji do zastosowania w robotach rehabilitacyjnych i ich wpływu na skuteczność terapii małych pacjentów.

Keywords: *Emotions, robot, rehabilitation, ASD, therapeutic interventions, machine learning, artificial intelligence*

Słowa kluczowe: *Emocje, robot, rehabilitacja, ASD, interwencje terapeutyczne, uczenie maszynowe, sztuczna inteligencja*

Introduction

Analyzing emotions during a therapeutic session for children with autism spectrum disorder (ASD) and similar conditions is extremely important, both for therapists and for the children themselves. Children with ASD often exhibit unique communication patterns and may have difficulty communicating their feelings using traditional verbal and nonverbal methods. Therefore, analyzing subtle nuances in facial expressions, voice pitch, speaking rate, and other auditory expressions can provide key information about their emotional state, needs, and reactions to various therapeutic stimulations.

Children can express emotions in various ways, for example, through facial grimaces or changes in the tone of voice, which may indicate anger, sadness, or joy. Using techniques of analyzing the patient's emotions, therapists can better understand what the child is trying to communicate and how they react to the environment and interventions. For example, monotonous speech may suggest disengagement or anxiety, while a high-pitched voice may indicate excitement or stress. Emotion analysis in the therapy of children with ASD is also helpful in monitoring the progress of treatment. Regular sessions during which emotional expression changes are analyzed may indicate the effectiveness of the therapeutic methods used or the need to modify them. Thanks to this, therapists can adapt their approach to each child's individual needs, which increases therapy's effectiveness.

In clinical practice, the ability to analyze the emotions expressed by a young patient is as essential as observation and diagnostic skills. Training in recognizing specific features of expressing feelings, as well as speech and sounds, should therefore be an integral part of the education of therapists working with children with autism spectrum disorder and similar disorders. Emotions can be read not only from the tone of voice or other sounds but also from facial expressions, which is particularly important in the therapy of children with autism spectrum disorders (ASD) and other diseases. Facial expression analysis provides therapists valuable information that may be difficult to obtain through verbal analysis alone.

Children with ASD often have difficulty using speech to express emotions or may not be able to fully interpret the body language of others, which means analyzing their facial expressions can help them understand their internal state. Therapists use facial expression techniques to identify subtle signals such as smiling, forehead wrinkles, and eye movements that may indicate feelings such as joy, anger, sadness, or surprise.

It is important to remember that children with ASD may express emotions in less typical ways than neurotypical children. For example, their smiles may be less visible or not occur in expected contexts. Therefore, professionals working with such children must be extremely sensitive to nuances in facial expressions and other forms of nonverbal communication.

Assistive technologies such as facial expression analysis software can supplement clinical observations by providing more objective data about children's emotional responses. Such tools can be handy when experienced observers miss subtle signals.

Integrating the analysis of facial expressions with the analysis of voice and other sounds creates a comprehensive approach to understanding the emotions and behaviors of children with ASD. Thanks to this, therapists can better adapt their work methods to maximize each child's emotional and social development.

Automation of the rehabilitation process, especially in the context of monitoring progress and proposing changes in therapy, is an important aspect that is gaining importance in the treatment of children with developmental disorders, such as ASD. Emotion analysis plays a key role here because it allows you to collect and analyze objective data about the child's emotional state and reactions, which is extremely important for the effectiveness of therapy.

Benefits of automating emotion analysis in rehabilitation: Objective assessment of progress: Automation allows for systematic tracking of subtle changes in a child's behavior and emotional expression, which may be difficult to assess only through human observation. Therapists can receive regular reports on the patient's progress thanks to technologies such as facial expression or voice modulation analysis.

Real-time therapy adjustments: Based on collected data, programs can propose modifications to the therapeutic approach tailored to the child's current needs. For example, if analysis indicates increased stress levels, therapy may be modified to increase relaxation elements.

Increased parental involvement: Automation enables parents to be better informed about their children's progress by providing detailed reports and analyses. Parents can better understand their children's needs and support their development more effectively in the home environment. Optimization of therapy time: Automation of processes allows for more effective use of therapeutic time. Therapists can focus on more complex aspects of therapy now that they have access to initial data analyzed automatically.

The use of technology in rehabilitation automation: Modern technologies such as machine learning and artificial intelligence (AI) play a crucial role in automating emotion analysis. These systems can be trained to recognize emotional and behavioral patterns specific to individual children. This makes it possible to monitor progress and detect potential problems or regressions early, allowing for a quick response and adjustment of the therapeutic plan. Automation of emotion analysis in the rehabilitation of children with developmental disorders brings many benefits, both in terms of therapeutic effectiveness and in terms of supporting parents in the therapy process. Advanced technologies allow for better adaptation of interventions to each child's needs, which is crucial for their development.

The use of a rehabilitation robot in the therapy of children with developmental disorders, such as ASD, is an innovative approach that can significantly enrich the therapeutic process. Robots, acting as interactive tools, engage children through play and offer precise and objective methods to assess and support their development.

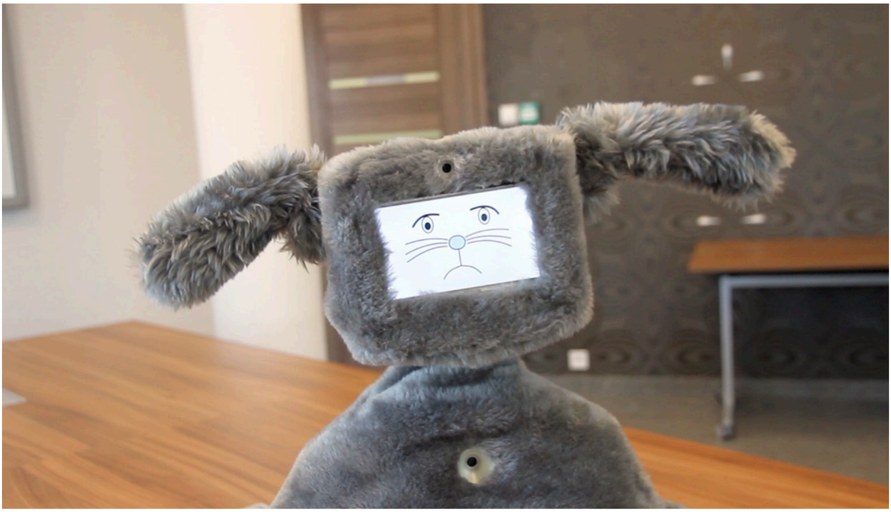
Robots can be programmed to carry out attractive games and activities for children, increasing their motivation and involvement in therapy. This allows children to participate better in sessions and derive more joy and satisfaction from them. Additionally, robots can be programmed to adjust their behavior depending on the child's response, allowing for a more personalized therapeutic approach. Thanks to advanced algorithms, robots can analyze a child's behavior, facial expressions, and voice, adapting their responses to maximize the effectiveness of therapy.

Robots equipped with sensors and cameras can constantly monitor and record a child's progress, giving therapists valuable insight into the effectiveness of therapy. This data can be used to tailor treatment programs and interventions further. Robots can also be used to teach social skills, such as emotion recognition and verbal and non-verbal communication. Children can learn these skills in a controlled but dynamic environment, often more effective than traditional methods (Schiavo et al. 2024). Interacting with a robot can be less stressful for children with ASD, who may sometimes feel anxious about interacting with other people. Robots offer repetition and predictability, essential to many children with ASD. Examples of robots include NAO, Pepper, PARO – a robot resembling a seal (Szymona et al. 2021, Puglisi et al. 2022), which is used mainly in sensory therapy, or a robot of our design (Niderla and Maciejewski 2021) used in research – show, how they can be used to reduce stress and improve social interactions. Robots during therapeutic sessions open new possibilities for the therapy of children with developmental disorders, offering innovative and effective support methods.

Data acquisition

The rehabilitation system consists of three components: the patient's device, in the presented example, is a robot; the rehabilitator's device, which is a computer with software that supervises and controls the robot's operation; and a machine learning module that processes the collected data, proposes actions appropriate to the current rehabilitation situation and training an artificial intelligence model. The robot has been designed to encourage children to interact, adopting a form inspired by nature. In this case, it draws on the appearance of a rabbit (Fig. 1), which makes it attractive to younger users. The robot's friendly, zoomorphic design, combining human and rabbit features, makes it charming and attracts children's attention. Its soft, gray fur-like coating enhances the sensory experience, making the robot visually and tactilely friendly. The robot comprises several critical structural modules, including the chassis, arms, neck, and head. Each is designed with the highest precision using 3D printing and laser cutting technology. This approach to making the device allows for flexible refinement of every detail and ensures sufficient durability and functionality. The robot chassis is equipped with an advanced drive system that allows smooth movement in various directions, which is achieved using DC motors and precise screw gear. Additionally, the robot uses optical sensors placed in its base, which allows it to avoid obstacles and prevent accidental falls. Robot arms, also created using 3D printing, have servo motors that enable complex movements such as lifting, tilting, or bending at the elbows. Thanks to this, the robot can grab and manipulate light objects, which can be especially useful when playing with children.

Figure 1. *Rehabilitation robot*



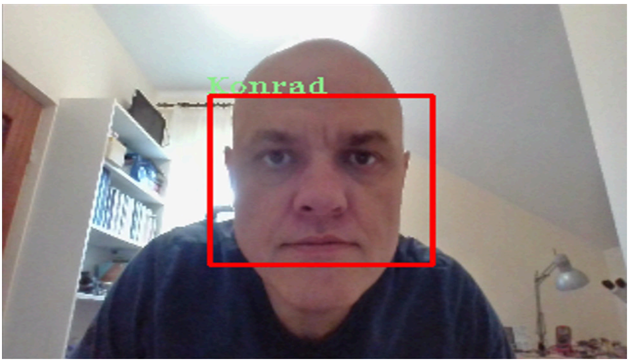
The robot's head is the center of communication and interaction. It contains speakers, a microphone, a camera, and a touch screen, which makes it highly functional. The microphone and camera record images and sound, allowing for advanced analysis of the child's voice and emotions. This, in turn, allows the robot to respond appropriately and adapt to the user's mood and needs, which significantly enriches interactions. The robot's head movements are designed to imitate human gestures such as nodding and tilting, which can be used to express emotions such as interest or sadness. The robot's ears, which can move independently, add another level of expression, as does the movable tail, which has a communication function, giving the whole thing a more dynamic and natural feel. The robot is not only a toy but also an educational tool that, thanks to its advanced technological functions and well-thought-out design, acts as an intermediary in the rehabilitation and education of children, supporting their emotional and cognitive development through interactive play and learning.

The rehabilitator's device is designed to receive data from the patient's device and enable interaction with the patient through various control elements. It contains a screen that displays real-time data and loopback recommendations generated by the machine learning system. The rehabilitator can use this device to modify therapeutic activities or adjust the response of the patient's device depending on the patient's current needs. The central element of this system is a machine learning module that processes the collected data using advanced algorithms to analyze and identify patterns. This system assesses the patient's emotional state and adjusts therapeutic activities to support the rehabilitation process best. The data processing unit analyzes information coming from the patient device while learning algorithms adapt the response of the patient device based on the patient's identified needs and responses. The system also has a dynamic feedback loop, where adjustments and recommendations are constantly sent between the machine learning system and the rehabilitator's device. This allows for ongoing adjustments to the therapy, which is crucial for the effectiveness of the treatment. In addition, the system can automatically adjust the patient's device responses to changing conditions and needs, increasing personalized therapy's effectiveness. This integrated rehabilitation system uses modern technologies and advanced machine learning to create a responsive and adaptive therapeutic environment that is tailored to the individual needs of patients, especially those with disorders such as autism. Through continuous data collection and analysis, this system enhances traditional therapy methods and provides rich data sets that can be used to refine treatments and therapies further.

Methods of analyzing emotions, voice, and sound

Emotion analysis based on camera images uses advanced technologies in pattern recognition, artificial intelligence, and image processing, which opens new possibilities in understanding human emotions. These tools allow you to precisely identify and analyze subtle signals that may indicate a person's various emotional states. One of the main methods is facial expression analysis, where the camera records facial movements and facial expressions, which are vital indicators of emotions. Machine learning and deep learning algorithms identify specific features such as smiling, frowning, and eye movements. Each of these changes in facial expression can suggest particular emotions, such as anger, sadness, or joy. Another important aspect is the analysis of gestures and body posture, which can be monitored using cameras. Hand gestures, body poses, and even how one moves can provide valuable information about a person's emotional state, which is particularly useful in social interactions or therapy (Landowska et al., 2022). Combining visual and audio data allows for even more accurate and comprehensive analyses. By integrating information from various sources, systems can better interpret complex emotional signals, which are vital in therapies, social interactions, and commercial applications, such as analyzing customer reactions. Each of these methods uses advanced algorithms and specialized software for data processing and analysis, and their effectiveness may depend on many factors, including the quality of the equipment, environmental conditions, and the specific nature of the application. Despite these challenges, emotion analysis technologies using cameras and microphones open new possibilities in understanding and interacting with the emotional world of patients. In the case of children with autism, who often encounter difficulties in verbal and non-verbal communication, modern technologies for analyzing emotions using cameras and microphones can significantly contribute to their therapy and rehabilitation. The rehabilitation robot, as a complex system for interacting with children, is equipped with a camera and a microphone that record facial expressions and gestures, as well as the tone and modulation of the voice. The collected data is analyzed to detect emotions such as satisfaction, frustration, or interest. In contrast, the microphones can analyze changes in the tone of voice, providing important information about the child's emotional state. With this data, we create a dataset linking the rehabilitation process with changes in patient moods to better support each child depending on their individual reactions and emotional needs. For example, if the system recognizes that a child is stressed, the system may suggest a calmer activity. A model may also encourage the child to interact when positive emotions are noticed, which helps build trust and emotional security. Emotion analysis, using technical devices such as cameras and microphones, can also be used to understand better the intentions and needs of children who may have difficulty expressing their thoughts and feelings clearly. This understanding is critical to personalizing therapy and tailoring interactions and subsequent rehabilitation steps, which can significantly improve the effectiveness of therapy. Integrating emotion analysis into therapy offers new opportunities for children with autism, helping them express and understand better and supporting their social and communication skills development in a more controlled, dynamic, and interactive environment.

Figure 2. *Face detection using the Haar Cascade method*



During the project, advanced technologies and algorithms were used to conduct a detailed analysis of data recorded during the robot's interactions with children, enabling precise understanding and response to their emotional and communication needs. The first stage uses face detection algorithms, vital in identifying and tracking the child's face in real time. The operation of popular and effective methods was tested here, such as Haar Cascades (Viola and Jones, 2001) or deep convolutional networks (CNN) (Mehendale, 2020), which can precisely recognize human faces in various lighting conditions and positions. Figure 2 shows face detection using the Haar Cascade method implemented in the OpenCV library based on the image captured by the computer's camera to determine when and where to focus the analysis of emotions and speech.

Figure 3. *Flowchart of the emotion detection process*

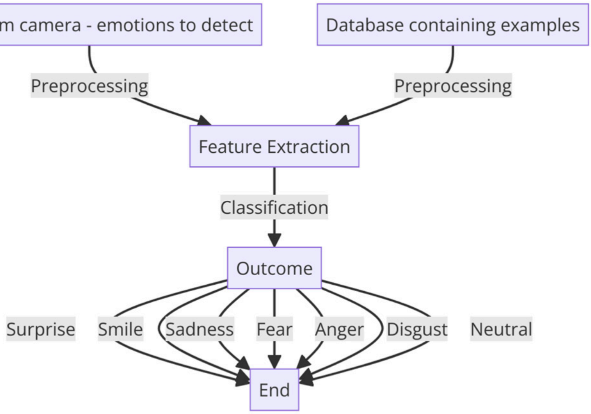
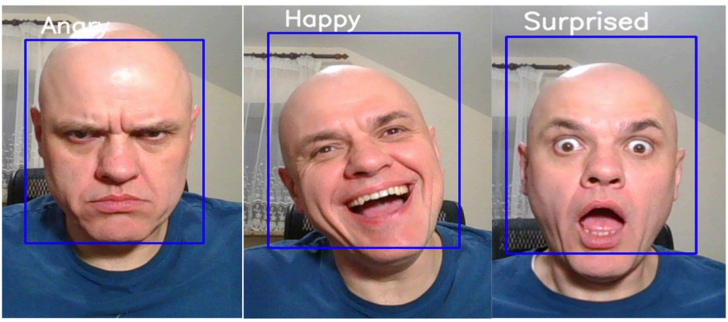


Figure 4. *Example of emotions recognized by the convolutional network model*

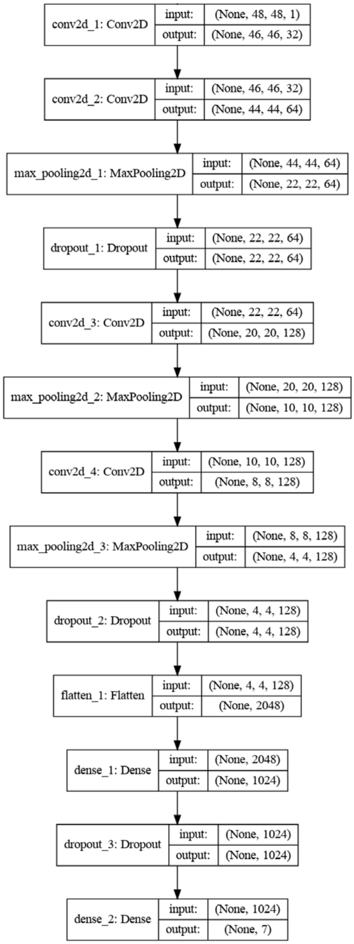


Recognizing emotions from human facial expressions is one of social communication's most complicated but crucial aspects (Cîrneanu et al., 2023). Facial expressions, as a natural and direct way for humans to express their feelings and intentions, play a central role in people's non-verbal exchange of information. Various facial expressions such as smile, sadness, anger, disgust, surprise, or fear are reflections of a person's internal state and means of communication that others can read. For example, a smile, one of the most accessible facial expressions, is characterized by raised corners of the mouth and raised cheeks, which can be further accentuated by tightening the lower eyelids. Sadness is expressed through furrowed brows, narrowed eyes, and twisted lips. Conversely, anger manifests itself through knitted and joined eyebrows, a horizontal wrinkle between them, tense lower eyelids, and a pursed mouth with the corners turned down. Enlarged nostrils and a forward chin may further accentuate some of these features. Disgust, in turn, can be signaled by knitted eyebrows and a wrinkled nose. Expressions such as surprise or shock are usually indicated by raised eyebrows, a raised upper lip, and a wrinkled nose, further emphasized by raised cheeks and wrinkles under the lower eyelids.

On the other hand, fear is characterized by raised and pursed eyebrows, often forming a straight line, with raised upper eyelids, tense lower eyelids, and a slightly open mouth with tense lips pulled back. Facial expression recognition includes critical steps such as feature extraction and classification. Feature extraction can be based on geometry (e.g., the shape of eyes, mouth, nose, eyebrows) or appearance, considering more detailed aspects of facial expression. Classification is the process by which the expressions mentioned above are categorized into smile, sadness, anger, disgust, surprise, and fear. Signals that can be read from the face are divided into static, slow, and fast. Static signals include skin color, facial shape, and bone structure. In contrast, slow signals appear over time and may consist of changes in facial appearance, such as permanent wrinkles or muscle tension. Fast signals are those associated with involuntary facial muscle movements, such as rapid changes in facial expression. Ultimately, work on facial expression recognition systems focused on three main stages: data pre-processing, feature extraction, and classification (Figure 3), which allows for adequate recognition and interpretation of human emotions based on facial observations.

The layer diagram of an example classifier constructed based on artificial neural networks is presented in Figure 5. Convolutional layers are the foundation of convolutional neural networks (CNNs), mainly for image processing. In this layer, each neuron is connected only to the local area of the input

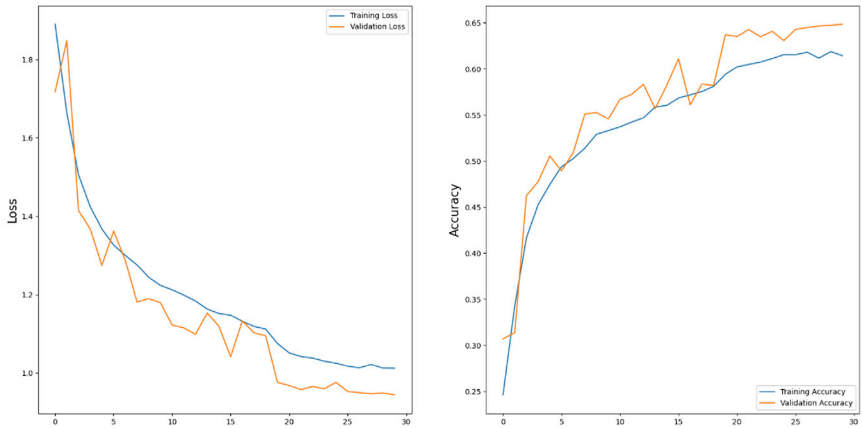
Figure 5. *Generated layer diagram of the convolutional neural network of the emotion recognition system*



data, which allows the detection of local spatial features such as edges, colors, and textures. Filters (small weight matrices) are moved across the entire input space (e.g., an image), creating feature maps that are the output of this layer. Each filter is responsible for detecting a specific type of feature at different levels of abstraction. Max pooling layers reduce the dimensions of the feature maps obtained in convolutional layers. By reducing the amount of data, max pooling helps reduce the computational complexity of the network and prevents overfitting. In practice, max pooling involves sliding a window (usually 2x2 or 3x3 pixels) through the feature map and selecting the maximum value in each window to represent a given area. The dropout regularization layer is used to prevent overfitting in neural networks. During training, dropout randomly *turns off* some neurons in the network, forcing the model to not rely on any individual neuron. This makes the model more robust and better generalizes it to new and unknown data. The flattened layer transforms a multidimensional feature map (e.g., from convolutional or pooling layers) into a one-dimensional vector. Such a transformation is necessary when subsequent layers in the network expect a one-dimensional vector of input data. Dense or fully connected layers are a standard element of many neural networks. Each neuron in the thick layer is connected to each neuron in the previous layer, which allows for the integration of information from earlier layers. Dense layers are often used at the end of neural networks for classification and regression, where the results from previous layers are transformed into the final model output, e.g., in the form of classes or predicted values.

Training a neural network involves adjusting the weights of individual neurons based on images of known classes (training set). Each training epoch adjusts the neuron weights, reducing errors and increasing the prediction accuracy (Figure 6). The trained network can extract features and classify previously unknown images.

Figure 6. *The plot of error and accuracy of the trained neural network*



Conclusions

Analyzing emotions in therapy sessions with children with autism spectrum disorder (ASD) is crucial to understanding their unique communication patterns and emotional states. These children often encounter difficulties in traditional forms of expressing feelings, which poses a challenge for them and therapists. Sound and facial expression analysis techniques, supported by modern technologies such as machine learning and artificial intelligence, allow for more accurate identification of subtle signals that may escape human observers. The introduction of rehabilitation robots and software for analyzing emotions based on speech and facial expressions opens up new possibilities in the individualization of therapy, adapting the treatment to the specific reactions and needs of the child. Thus, the use of these tools not only increases the effectiveness of treatment but also helps build more trusted therapeutic relationships, which is the foundation for adequate support for the development of children with ASD. Regular monitoring of progress and modifications in therapeutic approaches, supported by automation and data analysis, is essential to more effective and empathetic care for children with developmental disabilities. Voice and sound analysis in a rehabilitation robot is a separate issue that will constitute the scope of further work on using a rehabilitation robot. Speech-based emotion recognition uses various algorithms and machine learning models that analyze and classify audio data. In recent years, models based on attention mechanisms have gained popularity. They are now widely used in various sequential processing tasks, especially where mapping between two data formats is necessary. These models use previously predicted sequences and learn to map new sequences using an encoder-decoder approach, allowing for more efficient processing and better context understanding.

References

Bluestone, J. (2010). Materia autyzmu, przeł. M. Dąbrowska-Jędral, A. Klimek, Warszawa.

Cîrneanu, A.-L., Popescu, D., Iordache, D. (2023). New Trends in Emotion Recognition Using Image Analysis by Neural Networks, A Systematic Review. Sensors 23, 7092. https://doi.org/10.3390/s23167092

Landowska, A., Karpus, A., Zawadzka, T., Robins, B., Erol, Barkana, D., Kose, H., Zorcec, T., Cummins, N. (2022). Automatic Emotion Recognition in Children with Autism: A Systematic Literature Review. Sensors (Basel). 22(4):1649. doi: 10.3390/s22041649. PMID: 35214551; PMCID: PMC8875834.

Lord, C., Elsabbagh, M., Baird, G. et al. (2018). Autism spectrum disorder. The Lancet; 392 (10146): 508–520.

Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). SN Appl. Sci. 2, 446. https://doi.org/10.1007/s42452-020-2234-1

Niderla,K., Maciejewski, M. (2021). Construction of a zoomorphical robot for rehabilitation of autistic children Journal of Physics: Conference Series 1782, 012024 IOP Publishing doi:10.1088/1742-6596/1782/1/012024

Parsons, L., Cordier, R., Munro, N., et al. (2019). A Play-Based, Peer-Mediated Pragmatic Language Intervention for School-Aged Children on the Autism Spectrum: Predicting Who Benefits Most. J Autism Dev Disord. 10.

Puglisi, A., Caprì, T., Pignolo, L., Gismondo, S., Chilà, P., Minutoli, R., Marino, F., Failla, C., Arnao, A.A., Tartarisco, G., Cerasa, A., Pioggia, G., (2022). Social Humanoid Robots for Children with Autism Spectrum Disorders: A Review of Modalities, Indications, and Pitfalls. Children (Basel). 9(7):953. doi: 10.3390/children9070953

Schiavo, F., Campitiello, L., Todino, M.D., Di Tore, P.A. (2024). Educational Robots, Emotion Recognition and ASD: New Horizon in Special Education. Educ. Sci. 14, 258. https://doi.org/10.3390/educsci14030258

Smith, T. (2012). Evolution of Research on Interventions for Individuals with Autism Spectrum Disorder: Implications for Behavior Analysts. Behavior Analysis in Practice, 1 (35): 101–113.

Szymona, B., Maciejewski, M., Karpiński, R., Jonak, K., Radzikowska-Büchner, E., Niderla, K., Prokopia, A. (2021). Robot-assisted autism therapy (RAAT). Criteria and types of experi-ments using anthropomorphic and zoomorphic robots. Review of the research. Sensors 2021, 21, 3720. https://doi.org/10.3390/s21113720

Viola, P., Jones, M. (2021). Rapid Object Detection using a Boosted Cascade of Simple Fea-tures, Accepted conference on computer vision and pattern recognition 2001.