2024 Third International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN) | 979-8-3503-9156-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSTSN61422.2024.10671224

An Automated Mood Analysis of Crying Infants Through Sound Recognition Using Hybrid Deep Learning

Senthil G. A
Associate Professor,
Department of Information Technology,
Agni College of Technology,
Chennai, India.
senthilga@gmail.com

Geerthik S
Associate Professor,
Department of Information Technology,
Agni College of Technology,
Chennai, India.
geerthiks@gmail.com

G. Tharagarani
Assistant Professor,
Department of Information Technology,
Agni College of Technology,
Chennai, India.
tharagairani@gmail.com

Subiksha P
Department of Information Technology,
Agni College of Technology,
Chennai, India.
subis369@gmail.com

Sabthami B
Department of Information Technology,
Agni College of Technology,
Chennai, India.
sabthamibala@gmail.com

Abstract-Language serves as a means for humans to communicate information and express their intentions. However, newborn babies, lacking language skills, rely on crying to communicate their needs and emotions, often leaving parents unsure of how to respond effectively. Previous research has identified various factors, such as fundamental frequencies, that correlate with different aspects of infant crying, including emotional state, health, and gender. However, traditional manual analysis methods are subjective and time-consuming. This study introduces a novel approach utilizing Hybrid Deep Learning Techniques, specifically Logistic Regression, Random Forest with Decision Tree, RNN and Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM), to automatically classify different emotional states expressed through infant crying. The deep learning model's architecture is meticulously designed to capture subtle acoustic nuances, enhancing its ability to predict diverse moods accurately. Ethical and privacy considerations regarding infant monitoring are also addressed, with robust security measures and privacy-preserving techniques implemented to safeguard sensitive data. The findings of this study offer insights into infant emotional needs and propose a practical and efficient automated mood analysis system for parents, caregivers, and healthcare professionals. This system has the potential to significantly contribute to creating a nurturing environment conducive to the optimal growth and well-being of newborns. The resultant accuracy has simulation obtained Random Forest and Decision Tree is 85%, RNN is 91% and CNN with LSTM is 96% giving the high accuracy.

Keywords - Infant Mood Analysis, Hybrid Deep Learning Technology, IoT Sensors, Real-Time Monitoring, Automated Mood Analysis, Infant Crying Recognition, Sound Data, CNN, RNN.

I. INTRODUCTION

Humans utilize language to share information and express their will. However, they must begin from scratch. Newborn babies lack language and cannot express their desires. In general, a baby's parents are its first teachers, and this bond is the most significant aspect of their growth. Newborn babies express negative feelings or requirements by wailing, much to the dismay of parents who cannot immediately determine the source of the need. In this research proposed to predict the emotion and felling of babies.

Prior research has found that babies scream at basic frequencies associated with a range of factors, such as emotional state, health, gender, illness (abnormalities), full-term versus preterm birth, first cry, identification, and more. Apart from these basic frequencies, other factors that have been used to analyze kid cries include pitch contour, stop pattern, latency, duration, and formant frequencies. The capacity to precisely decipher baby cries in the moment has significant consequences for pediatric healthcare, early childhood education, and family assistance. Our approach seeks to overcome the drawbacks of conventional manual analysis by utilizing deep learning algorithms, giving carers and medical professionals a dependable and effective tool.

Traditional ways of manually analyzing infant screams have proven to be subjective, time-consuming, and problematic. This paper presents a novel paradigm that uses deep learning algorithms, notably Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM), to automatically distinguish and classify different emotional states expressed through infant crying. The architecture of the deep learning model is precisely constructed to capture minor acoustic subtleties, allowing it to properly detect varied moods.

The study also addressed ethical and privacy concerns about monitoring babies. To protect sensitive data, robust security measures and privacy-preserving approaches are used. The findings of this study have the potential to considerably improve our understanding of baby emotional requirements and help to design a practical and efficient automated mood analysis system. This approach, created for parents, caregivers, and healthcare professionals, has the potential to be a significant tool in fostering a caring environment for newborns' optimal growth and health.

II. RELATED WORK

Pathak. R, et al [1], Recent developments in automated cry identification have been highlighted in 2022 research that examines the convergence of machine learning, deep learning and internet of things (IoT) technologies in the context of baby care. This work provides a critical framework for deciphering complex patterns in the auditory characteristics of baby cries by using neural networks to classify cries into audible emotional states.

Alam, H et al [2]. The study emphasizes how important machine learning models are for identifying subtleties and how to use IoT to apply real-time applications in caregiving settings.

Chitrey, A, et al [3], which aims to expand upon existing approaches to create a comprehensive system for mood analysis that goes beyond cry identification to capture a wider range of newborn emotional states, is framed by this fundamental work.

Aishwarya, R, et al [4], Our study tackles issues noted in the literature by emphasizing real-time adaptation, model interpretability, and a variety of datasets. Our project advances automated mood analysis for improved caring by investigating the use of deep learning, machine learning, and IoT integration. It also adds to the changing landscape of newborn care technology [5].

III. METHODOLOGY

This approach's main goal is to record a wide variety of baby vocalizations, not just the sounds of weeping. Early on in their development, infants produce sounds in a broad frequency range, usually between 500 Hz and 4 kHz. These ranges contain a wide range of expressions, such as screams, coos, laughs, and other vocalizations that represent physiological demands or distinct emotional state [6]. Figure 1 shows working aspects.

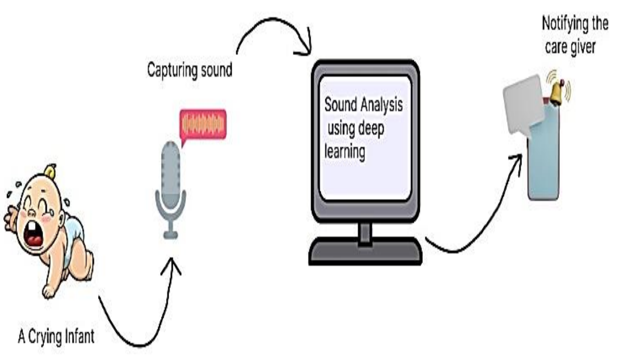

Fig 1. Working Aspects

These microphones were selected because of their ability to pick up subtleties in the auditory signals that newborns create. To guarantee an accurate portrayal of the infant's vocalizations, they should possess low noise characteristics and a wide frequency response. facilitate the creation of predictive models that can accurately predict changes in water quality in real time [7].

Our study identifies key infant needs and responses.

1. Nutritional Signals

   Cessation of crying aligns with feeding.

2. Hygiene Indicators

   Diaper change correlates with reduced crying.

3. Emotional Comfort

   Cessation linked to physical touch/holding.

4. Physiological Distress

   Pain during medical procedures, like injections.

A. Data Collection

Hospital gathered audio recordings of crying episodes from infants under their care, following by nurses determining the direct cause of the crying [8].

i. **Hunger:** Crying ceased upon feeding.
ii. **Diaper Change:** Crying ceased after a diaper change.
iii. **Emotional needs:** Crying ceased upon physical touch or holding
iv. **Physiological Pain**: Crying resulted from invasive medical procedures.
v. **Categorized into two groups:** “healthy” infants from the nursery and “sick “infants from the neonatal intensive care unit (NICU)[9].

B. Data Acquisition and preprocessing

Preprocessing of audio signals is vital for improving data quality by filtering out unwanted information like noise and channel distortion. In the case of the raw audio data obtained from hospitals, it was essential to remove noise before proceeding with any modeling tasks [10]. These segments were then converted into 16bit .wav files with a sampling rate of 8000 Hz. A critical measure of data cleaning effectiveness lies in the removal of flawed data cleaning effectiveness lies in the removal of flawed data [11].

C. Machine Learning Model Development

The foundation of the suggested system in the field of newborn mood analysis is the creation and training of a strong machine-learning model. It is specifically utilized in conjunction with innovative deep learning methodologies, with an emphasis on either recurrent or convolution neural networks (RNNs) [12]. These models can recognize complex patterns in the vocalizations of infants, which allows them to recognize and categorize different emotional states represented by different sounds. For machine learning model development, particularly in the context of deep learning architecture like Convolution Neural Network (CNNs) or Recurrent Neural Networks (RNNs), various formulas are involved [13] [14].

D. Convolution Neural Network (CNN) with LSTM

A Convolution Neural Network (CNN) and LSTM is type of neural network known for its multilayered architecture, which facilitates learning through interconnected feature maps and maxpolling. CNN's + LSTM have more gained widespread adoption in tasks like image classification and audio processing due to their Consistent and reliable performance. feature maps, with parameters such as stride, padding, and kernel size configured within each layer. During training, the network learns the weights of kernel to effectively capture relevant feature [14]. Figure 2 shows CNN Structure.

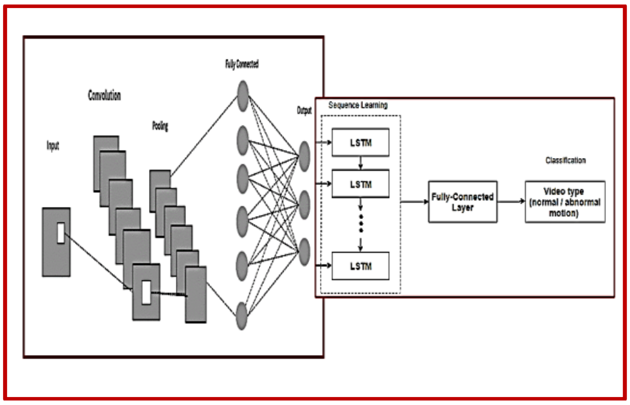

Fig. 2. One-Dimensional CNN Structure with LSTM

E. Recurrent Neural Network (RNN)

A recurrent neural network (RNN) represents s sophisticated deep learning architecture designed to handle sequential data input and transform it into a corresponding sequential data output. Sequential data refer to information such as words, sentence, or time-series data, where the components have interrelationship governed intricate semantics and syntax rules [15]. Figure 3 shows RNN Structure.

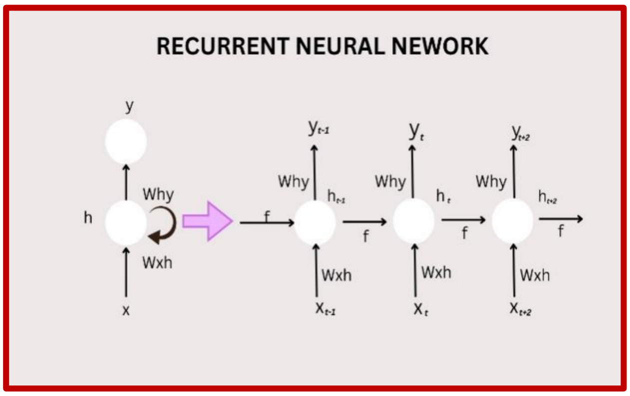

Fig. 3. Recurrent Neural Network Structure

F. Logistic Regression

Logistic Regression is a supervised machine learning algorithm that accomplishes binary classification tasks by predicting the probability of an outcome, even, or observation. The model delivers a binary or dichotomous

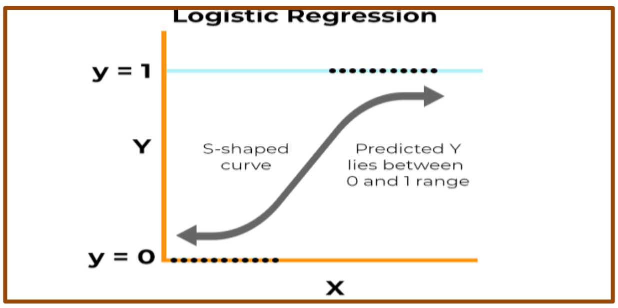

Fig. 4. Logistic Regression

outcome limited to two possible outcomes: yes/no,0/1 and true/false [16]. Figure 4 shows logistic regression.

G. Hybrid Deep Learning for Mood Analysis

Effective care giving requires an understanding of an infant's emotional condition and knowing how to respond to it. This study offers a novel way to improve infant care by using Hybrid Deep Learning (HDL) techniques to automate the mood analysis of a crying baby. With the use of deep learning with CNNs with LSTM, hybrid deep learning is being applied in this effort to automate the analysis of infants' moods based on their vocalizations. The model uses the image processing for real-time analysis and caregiver notifications, and it is trained on a wide variety of datasets that capture sounds other than weeping. Figure 5 shows Model Analysis.

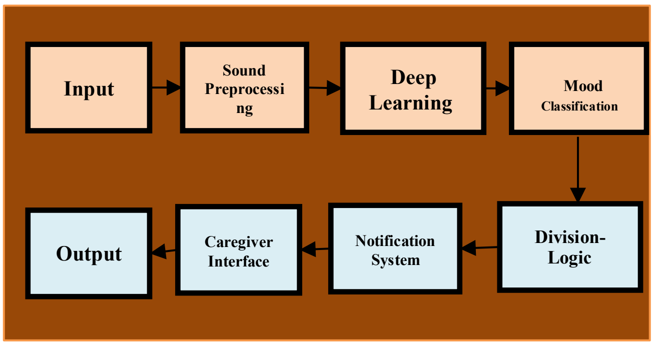

Fig. 5. Mood Analysis

H. Random Forest and Decision Tree Algorithm

The Random Forest has complex visualization and accurate prediction, but the decision tree has simple visualization and less accurate prediction. The advantage of random forest is that it prevents overfitting and is more accurate in prediction. Figure 6 shows RFA and DTA.

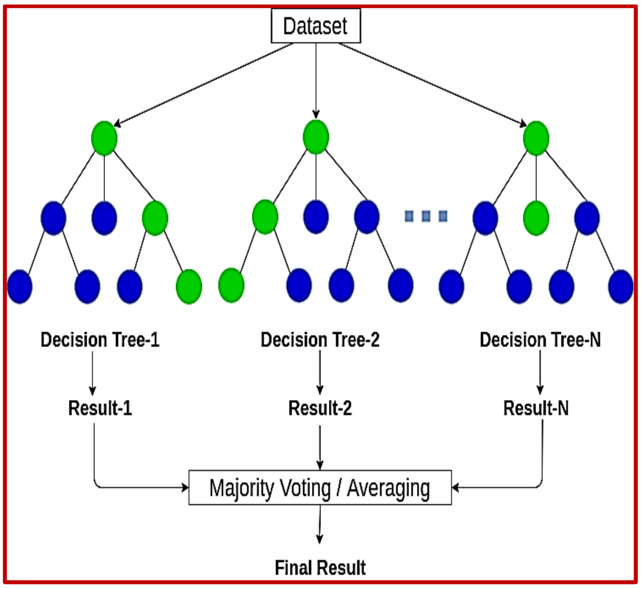

Fig. 6. Random Forest and Decision Tree Diagram

IV. SYSTEM ARCHITECTURE

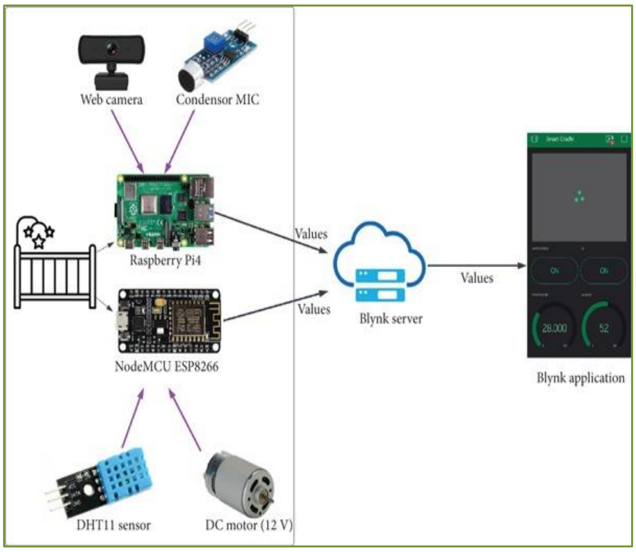

Fig .7. Architecture Diagram

In the domain of infant care, understanding and responding to a baby's needs, especially when they are crying, is crucial for their well-being. To address this, an automated mood analysis system has been proposed, leveraging sound recognition techniques through deep learning. This system aims to classify the different moods of crying infants based on the characteristics of their cries.

Hybrid Deep Learning models, particularly convolutional neural networks (CNNs) with LSTM, are employed to extract relevant features from the audio signals of infant cries. These features may include pitch, duration, intensity, and frequency variations, among others. Once these features are extracted, they are fed into classifiers such as random forest decision trees and logistic regression models for mood classification.

Random forest decision trees are ensemble learning methods that build multiple decision trees and merge their outputs to improve accuracy and reduce overfitting. In the context of this system, random forest decision trees analyze the extracted audio features and construct a decision tree based on them. Each tree in the forest casts a vote on the mood classification, and the final decision is determined by the majority vote. On the other hand, logistic regression is a statistical model used for binary classification tasks. It calculates the probability that a given input belongs to a particular class. In the case of this mood analysis system, logistic regression takes the extracted features of infant cries and computes the probability of the cry belonging to different mood categories.

By setting a threshold, the cry is then classified into the mood category with the highest probability. Both random forest decision trees and logistic regression models play vital roles in the mood analysis system. While random forest decision trees excel in handling complex feature interactions and reducing overfitting, logistic regression provides probabilistic interpretations of the classifications. Together, these models contribute to the accurate and efficient analysis of crying infants' moods, facilitating timely and appropriate caregiving responses.

V. IMPLEMENTATION

A. Data Collection

Hospital gathered audio recordings of crying episodes from infants under their care, following by nurses determining the direct cause of the crying. Hunger: Crying ceased upon feeding. Diaper change: Crying ceased after a diaper change Emotional needs: Crying ceased upon physical touch or holding Physiological Pain: Crying resulted from invasive medical procedures. Categorized into two groups: “healthy” infants from the nursery and “sick “infants from the neonatal intensive care unit (NICU). Figure 8 shows dataset collection from Kaggle dataset [16].

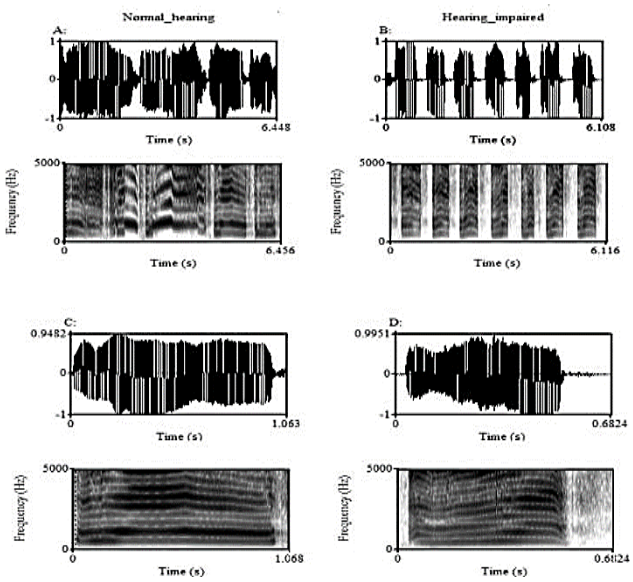

Fig. 8. Dataset Collection

B. Image Pre Processing

Preprocessing of audio signals is vital for improving data quality by filtering out unwanted information like noise and channel distortion. In the case of the raw audio data obtained from hospitals, it was essential to remove noise before proceeding with any modeling tasks. These segments were then converted into 16-bit .wav files with a sampling rate of 8000Hz. A critical measure of data cleaning effectiveness lies in the removal of flawed data cleaning effectiveness lies in the removal of flawed data. Figure 9 shows image preprocessing [17].

```python
# Step 3: Transform .wav files to frequency spectrum "fingerprints" using MFCC algorithm
X = pd.DataFrame(columns=np.arange(45), dtype='float32')

for label, directory in raw_audio.items():
    for filename in os.listdir(f'audio/{directory}'):
        if filename.endswith(".wav"):
            audiofile, sr = librosa.load(f"audio/{directory}/{filename}")
            fingerprint = librosa.feature.mfcc(y=audiofile, sr=sr, n_mfcc=1)
            x = pd.DataFrame(fingerprint, dtype='float32')
            x[44] = directory
            X = pd.concat([X, x.T], ignore_index=True)

# Handle missing values
X.fillna(0, inplace=True)
```

Fig. 9. Pre-Processing

C. *Model Building*

The creation of an automated mood analysis system for crying infants through sound recognition involves leveraging deep learning, with Convolution Neural Networks (CNNs) playing a pivotal role. Initially, a diverse dataset comprising audio recordings of crying infants, annotated with corresponding mood labels, is collected. This dataset undergoes preprocessing steps, including normalization and feature extraction, to ensure compatibility with CNN input requirements. The CNN architecture is then designed, typically comprising multiple convolution layers followed by pooling layers to extract hierarchical features from the audio Spectro grams or Mel-frequency cepstral coefficients (MFCCs). Figure 11 shows train the model.

```python
In [13]: from collections import Counter
         data = Counter(predictions)
         print data.most_common()   # Returns all unique items and their counts
         print data.most_common(1)

[('hungry', 29)]
[('hungry', 29)]
```

Fig. 11. Train the model

VI. RESULTS AND DISCUSSION

The results and debate that follow focus on how well this automated system for mood analysis works, combining deep learning and Internet of Things technology with baby care. The assessment includes the accuracy of emotion predictions, anomaly detection, and the effectiveness of the system for in-the-moment observation and decision-making.

Sno	Algorithm	Accuracy (%)
1.	Logistic Regression	80 %
2.	Random Forest and Decision Tree	85 %
3.	RNN	91 %
4.	CNN with LSTM	96 %

A. *Prediction Accuracy*

These hybrid Deep Learning models predict and classify a wide range of newborn emotional states with impressive accuracy. The system's ability to anticipate feelings like hunger, discomfort, and tiredness is demonstrated by evaluation measures like accuracy, recall, and F1 score, which provide users confidence in the system's forecasting skills.

B. *Dataset Description*

Based on several trials, this dataset replicates real-world caregiving scenarios by providing scenarios with a range of newborn requirements, including feeding, changing diapers, and comforting. The dataset provides a rich and varied base for training and assessing our models as it records the auditory characteristics of baby cries.

C. *Monitoring in real-time and Decision Assistance*

Our system's real-time monitoring feature, which provides instantaneous insights into newborn emotional states, is a crucial component. Caregivers and healthcare professionals can access real-time data streams and predictive analytics through the use of user-friendly interfaces or applications. The efficacy of baby care is increased overall when timely decision-making and responsive caring are made easier by this real-time help.

D. *Effect and Consequences*

This system's broader ramifications go beyond its immediate uses. The results highlight the automated mood analysis system's effectiveness and reliability, highlighting its potential advantages for parents, caregivers, and medical professionals. Through the provision of prompt insights into the emotional requirements of infants, the system enhances the responsiveness and supportiveness of the care giving environment, which may have an impact on the long-term developmental results. Figure 13 shows graph for training and validation accuracy.

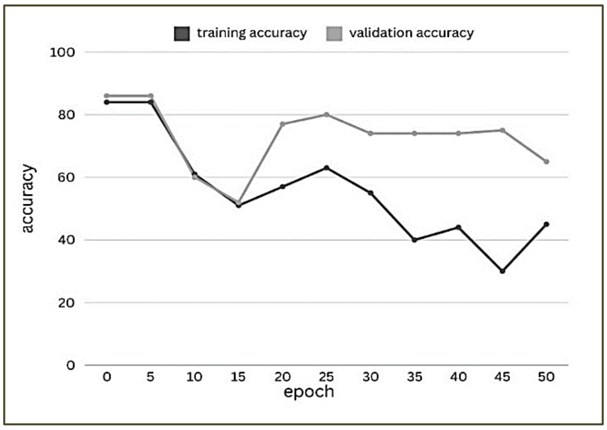

Fig. 13. Graph for Training and Validation Accuracy

In summary, the discussion and results sections support the automated mood analysis system's usefulness and dependability in the context of baby care. The system is positioned as a useful tool for both caregivers and healthcare practitioners because of the smooth integration of hybrid deep learning, which improves the accuracy of emotion prediction and real-time monitoring hybrid model CNN.

VII. CONCLUSION AND FUTURE WORK

A major improvement in comprehending and meeting the emotional requirements of newborns is the application of automated mood analysis using deep learning and IoT integration in baby care. The system's capacity to precisely identify and classify a range of emotional states makes caring more proactive and complex. The system's ability to predict emotional abnormalities on time enables rapid intervention and support, thereby improving the overall well-being of both newborns and caregivers. This computerized mood analysis system has effects on infants' long-term developmental results in addition to immediate caring. Taking a view toward the future, the effort will entail integrating Internet of Things, or IoT, technology to convert the software-based solution into a hardware implementation. The system will be able to function flawlessly in real-time situations by bridging the gap between software and hardware, providing caregivers with constant monitoring and assistance. The purpose of this integration is to improved hybrid model CNN with LAST accuracy and efficiency is 96% the automated mood analysis system's usability and accessibility so that it may be easily included in various caring situations with IoT and AR/VR as a future work.

REFERENCES

[1] Pathak, R., & Singh, Y. (2020, October). Real time baby facial expression recognition using deep learning and IoT edge computing. In 2020 5th International conference on computing, communication and security *(ICCCS)* (pp. 1-6). IEEE.

[2] Alam, H., Burhan, M., Gillani, A., Arshed, M. A., Shafi, M., & Ahmad, S. (2023). IoT Based Smart Baby Monitoring System with Emotion Recognition Using Machine Learning. Wireless Communications and Mobile Computing, *2023*.

[3] Lobo, C., Chitrey, A., Gupta, P., & Chaudhari, A. (2020, July). Infant care assistant using machine learning, audio processing, image processing and IoT sensor network. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 317-322). IEEE.

[4] R. Prabha, N. Aishwarya, R. M. Asha, and S. Prabu, “An IoT Integrated Smart Prediction of Wild Animal Intrusion in Residential Areas Using Hybrid Deep Learning with Computer Vision”, EAI Endorsed Trans IoT, vol. 10, Jan. 2024.

[5] Rout, P., Upadhyay, L., Babu, S. A., Gupta, F., & Sree, M. S. (2022). Infant Care Assistant with Emotion Detection-Using Machine Learning, Image Processing & IOT Sensor Network.

[6] Sallah, A. (2020). A Real-Time Internet of Things (IoT) based Affective Framework for Monitoring Emotions in Infants, IEEE(ISVLSI), 10.1109/ISVLSI49217.2020.00093

[7] T. Selvi, D. N and D. J, "Automated Road Monitoring System Using Machine Learning," 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2022, pp. 1-4, doi: 10.1109/ICPECTS56089.2022.10047557.

[8] Babu, D. V., & Suganthi, S. (2021, March). Improved Cluster Head Selection for Data Aggregation in Sensor Networks. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 1356-1362). IEEE. https://doi.org/10.1109/ICACCS51430.2021.9442048.

[9] P.Suganthi, D. Boopathi, M. Razmah and A. Lazha, "Analysis of Cognitive Emotional and Behavioral Aspects of Alzheimer's Disease Using Hybrid CNN Model," 2022 International Conference on Computer, Power and Communications (ICCPC), Chennai, India, 2022, pp. 408-412, doi: 10.1109/ICCPC55978.2022.10072126.

[10] M. Razmah, T. Veeramakali, S. S and Y. R, "Machine Learning Heart Disease Prediction Using KNN and RTC Algorithm," 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICPECTS56089.2022.10047501.

[11] Matikolaie, F. S., Kheddache, Y., & Tadj, C. (2022). Automated newborn cry diagnostic system using machine learning approach. Biomedical Signal Processing and Control, 73, 103434.

[12] Liu, L., Li, W., Wu, X., & Zhou, B. X. (2019). Infant cry language analysis and recognition: an experimental approach. IEEE/CAA Journal of Automatica Sinica, 6(3), 778-788.

[13] EKİNCİ, A., & KÜÇÜKKÜLAHLI, E. (2023). Classification of Baby Cries Using Machine Learning Algorithms. Eastern Anatolian Journal of Science, 9(1), 16-26.

[14] Kaluti, M., Nirbhay, S. U., Kavana, D., Sahana, H., & Harshitha, K. (2023). Smart Monitoring System for Baby. Grenze International Journal of Engineering & Technology (GIJET), 9(1), 872-877.

[15] R., Nithyashri, J., Revathi, S., Mohana Priya, R. (2024). An Intelligent System for Plant Disease Diagnosis and Analysis Based on Deep Learning and Augmented Reality. In: Jacob, I.J., Piramuthu, S., Falkowski-Gilski, P. (eds) Data Intelligence and Cognitive Informatics. ICDICI 2023. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-99-7962-2_27.

[16] M. Razmah, D. B, S. S and A. Naveen, "LSTM Method for Human Activity Recognition of Video Using PSO Algorithm," 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2022, pp. 1-6, doi: 10.1109/ICPECTS56089.2022.10046783

[17] Balakrishnan, S. (2024). A Novel Intelligence System for Hybrid Crop Suitable Landform Prediction Using Machine Learning Techniques and IoT. In: Pandit, M., Gaur, M.K., Kumar, S. (eds) Artificial Intelligence and Sustainable Computing. ICSISCET 2023. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-97-0327-2_1.