Egyptian Journal of Artificial Intelligence
ISSN (Online): 2786-0213, ISSN (Print): 2786-0205
Volume 3, Issue 2, 2024, PP. 1-10

AI-Powered ICare Infant Cry Analysis and Speech Therapy

Mahmoud Y. Shams[1], Nada Nasr[1], Heba Abdelaziz[1], Abeer Elmorshdy[1], Weaam Elsayed[1], Youmna Elghanam[1], Abdelrahman Shady[1], Abdelrahman Elbadawy[1], Osama Elmoghazy [1] , Ahmed Muqawi[1], Muhammed Khalifa[1], Abdelkareem Elkhateb[1], Ahmed Ghaly[1]

[1]Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh 33516, Egypt

mahmoud.yasin@ai.kfs.edu.eg, nadanasr370@gmail.com, hebauwk25800@gmail.com, byralmrshdy41@gmail.com, weaam.ai_0482@ai.kfs.edu.eg, youmnaelghanam226@gmail.com, abdooshady@gmail.com, gekaa1052@gmail.com, osamazxzx64@gmail.com, ahmedmuqawi@gmail.com, mohamed.ai_0400@ai.kfs.edu.eg, kareem01095134688@gmail.com, ahmedghaly0767@gmail.com

Abstract: This paper introduces ICare, a mobile application powered by artificial intelligence designed to support parents in raising their children, specifically those aged one month to one year. The app utilizes advanced AI to analyze baby cries and determine the cause, aiding parents in understanding their child's needs. Additionally, ICare offers speech therapy features with customized exercises to enhance children's language and communication skills. A chatbot provides immediate assistance in emergencies, and the app includes health and medical information for parents to stay informed about diseases and symptoms. The application is highly efficient, has been thoroughly tested, and is regularly updated. With an overall accuracy of 74.00%, the model excels in predicting certain classes, such as "hungry," though it needs precision improvement for classes like "tired." These results underscore the app's effectiveness in helping parents understand and respond to their baby's cries.

Keywords: Childcare, Artificial Intelligence, Emotion Recognition, Sentence Transformers, Voice Recognition, Speech Therapy, Chatbot, Mobile Application.

1. Introduction

It's critical to provide parents with the resources and tools they need to raise their kids well in the quickly changing world of today. It might be tough for parents to meet their child's demands and assure their well-being because of the many difficulties and uncertainties that come with modern globalization. This article presents a novel approach to improving the ICare mobile application. ICare's mission is to empower parents, especially those whose kids are between one month and one year old, by providing a comprehensive platform with interactive features, monitoring tools, and priceless advice to help them throughout the parenting path. Numerous important facets of childcare are covered by ICare. With the use of state-of-the-art artificial intelligence, the app can identify the root causes of a baby's screams and analyze them, assisting parents in understanding their child's feelings and the need for prompt and appropriate care. Furthermore, ICare offers personalized speech therapy activities to assist young children's language and communication development. The incorporation of these AI-powered functionalities marks a noteworthy progression in augmenting human experience via intelligent technology [1]. Additionally, ICare has a feature called Intelligent Chatbot Support. This clever AI chatbot provides real-time responses and support, making it an excellent choice for parental guidance. Additionally, the app creates a helpful community where caregivers may talk, exchange information, and get guidance. Through helpful hints, emotional support, and cooperative problem-solving, this community component improves caring abilities and gives caregivers a sense of empowerment and camaraderie. Moreover, ICare offers medical

information to parents to help them understand the health of their children. With the help of these extensive features, ICare makes sure that parents have all they need to confidently and resiliently handle the challenges of contemporary parenthood.

1.1 Problem Statement

In today's fast-paced and globalized world, parents face numerous challenges and uncertainties in meeting their child's needs and ensuring their well-being. The lack of accessible, reliable, and comprehensive tools makes it difficult for parents to effectively manage childcare, leading to stress and potential developmental issues for their children.

1.2 Objectives

- To develop a comprehensive mobile application, ICare, that supports parents in raising their children aged one month to one year.
- To utilize advanced artificial intelligence techniques for analyzing and predicting the causes of a baby's cries.
- To provide personalized speech therapy exercises to aid in the language and communication development of young children.
- To offer real-time support and guidance through an intelligent chatbot.
- To create a supportive community for caregivers to share information and receive emotional support.
- To supply parents with reliable health and medical information to better understand and care for their children's health.

1.3 Significance of the Study

This study is significant as it addresses a critical need for modern parents to have effective tools and resources for childcare. By leveraging AI technology, the ICare app provides innovative solutions that enhance the parenting experience, improve child development outcomes, and reduce parental stress. The study's findings contribute to the growing field of AI applications in healthcare and child development, showcasing the potential of technology to improve daily life.

1.4 Research Gap

Comprehensive AI-driven mobile apps that provide personalized speech therapy exercises, analysis of a baby's screams in advance, and real-time chatbot support for parents are still unobtainable in the market, despite technological breakthroughs. Existing solutions are usually piecemeal, covering either requirement for monitoring or needs for development, but rarely at the same time. This disparity emphasizes the need for a comprehensive platform such as ICare, which integrates these vital components to offer parents all-encompassing assistance. ICare seeks to close this gap by combining cutting-edge AI capabilities with useful solutions, providing parents with a one-stop shop for all their daycare needs. [2,3].

1.5 Motivation

The creation of ICare was motivated by the difficulties that parents have in the modern, international, fast-paced world. It might be difficult to understand a baby's demands and to ensure their wellbeing, therefore it's critical to have cutting-edge technologies that offer helpful insights and real-time support. To meet these needs, ICare provides a comprehensive platform that combines quick chatbot support with advanced AI-driven features like cry analysis and customized speech therapy activities. This guarantees that parents get timely and helpful advice, enabling them to more confidently and easily handle the challenges of contemporary parenting [4,5].

1.6 Main Contribution

The main contributions of this paper are:

- Development of the ICare app, an AI-powered mobile application tailored for parents with infants.
- Implementation of advanced AI algorithms to analyze and predict the reasons behind a baby's cries.
- Integration of personalized speech therapy exercises within the app.
- Deployment of an intelligent chatbot for real-time parental support and emergency assistance.
- Creation of a community platform within the app for caregivers to interact, share, and support each other.
- Provision of comprehensive health and medical information to educate and empower parents.

The paper is organized as follows: Section 2 covers related work, Section 3 details the research methodology, Section 4 presents the evaluation results, and Section 5 concludes with future work.

2. Related works

Work in the childcare industry has developed using a variety of algorithms and techniques. This section contains a few older publications that have furthered this field of study. Cry translation has been approached using a variety of methods and a related idea. An example of this may be seen in the work of a group at Yunlin University, which used a combination of convolutional neural networks to determine whether the spectrogram it was given was a baby's cry and then classified the purpose behind that cry into one of four categories [6]. Another group, Universitas Indonesia, linked convolutional and recurrent neural networks to create a more efficient system where the recurrent neural network learns from the functions obtained by the convolutional neural network [7]. Convolutional neural networks were also the subject of research at Georgia State University, which sought to enhance them by applying a multi-level convolutional neural network with a hybrid characteristic set and previous information [8]. Using their audio record facts, a team at Koç University fed the audio alerts from spectrograms into a capsule community that was evaluated directly using typical convolutional neural networks to refute these theories [9]. Researchers were unable to create a laptop application that even came close to wearing out discussions, so they gave up on the outdated methods of fusing linguistics and records. Rather, they used statistical models, even though early machine learning challenges these patterns and eventually results in the development of several high-level statistical natural language processing applications [10]. Two earlier studies that have addressed this issue are Language Development Screening for youngsters [11] and Systematic Analysis of Language Transcripts (SALT) [12], which were created and utilized in English-speaking nations to assist youngsters in speaking correctly. Nevertheless, the I Care service is utilized in both Arabic- and English-speaking nations, therefore both versions are used in these countries. In this study, we investigated how sentence-level embeddings can be used by speech therapy apps to support children's language and communication development. These embeddings support a variety of tasks, including sentence building, vocabulary enrichment, and pronunciation practice, by capturing semantic similarities between sentences. One such model, called Sentence Transformer "paraphrase-multilingual-MiniLM-L12-LM" [13], aims to generate high-quality paraphrases in many languages. This method uses a transformer-based architecture to encode input words into dense vector representations, allowing for fast comparisons and semantic matching. By creating interactive, individualized lessons based on the Sentence Transformer paradigm, speech therapy applications can assist children with their vocabulary, pronunciation, and sentence structure.

3. Research Methodology

The introduction of AI into childcare signifies a dramatic change in the ways that parents can handle the challenges involved in raising children. ICare hopes to deliver useful information and direction in addition to real-time insights into a child's emotional state by utilizing the power of machine learning algorithms [14]. This strategy goes beyond the project's immediate goals and advances a developing field that aims to improve human experience via intelligent AI integration. We will go into more detail about this system's implementation and examine the sociological implications as well as the technical nuances of implementing AI in childcare in the parts that follow. We will talk about three important topics: chatbot assistance, speech therapy, and emotion identification [15] using weeping sound analysis.

3.1 Emotion Recognition through Crying Sounds

In toddler communication, crying is a basic aural representation that captures a range of norms and emotional states. Fostering strong, organic connections requires an understanding of these auditory cues and an ability to respond to them. Using the "Donate a Cry" archival collection, we conduct a focused inquiry to investigate the operationalization of emotional deduction in relation to baby cries. This method categorizes and articulates the various emotional states expressed by crying using algorithmic learning paradigms, feature extraction, and audio preprocessing. The rhythmic intensity envelope of a normal newborn cry is seen in Fig. 1, with cyclic changes and a high pitch of approximately 500 hertz. One distinguishing characteristic of a toddler's cry is its periodic pattern, which is marked by alternating sobs and inspirations. Our goal is to accurately comprehend and respond to the emotional situations conveyed by baby cries by examining these sound patterns [16].

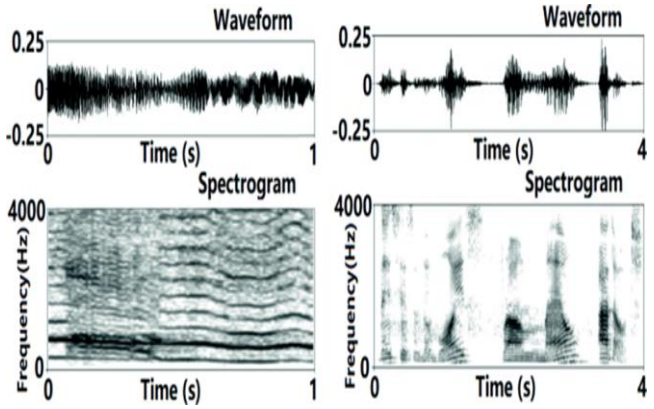

Fig. 1. The waveform of the voice of crying.

3.1.1 Dataset

The "Donate a Cry" dataset serves as the basis for training the emotion recognition algorithm [17]. This painstakingly selected dataset was created especially for this project and includes a broad variety of sobbing noises linked to various emotional states, such as weariness, discomfort, hunger, and stomach pain. This dataset's complexity and diversity allow the model to learn and generalize patterns well, enabling reliable emotion recognition. The system can correctly recognize and react to the different emotional indicators conveyed by a baby's screams by utilizing this extensive dataset, giving parents timely and relevant help.

3.1.2. Audio Preprocessing

To maintain consistency and alignment with future analytical processes, the raw audio recordings must go through a preprocessing phase before feature extraction can start. The raw audio recordings are first transformed into a format that is appropriate for analysis. After that, the audio waveform is analyzed to extract important properties, such as duration, frequency, and amplitude. The normalized input produced by this kind of preprocessing is essential for improving the computational model's capacity to recognize and decipher emotional subtleties in the audio signals. We carefully listened to each recording in the dataset during the audio preprocessing stage, personally eliminating any unnecessary information and null voices. With this meticulous preparation, the dataset is guaranteed to be clean and well-suited for the AI model's accurate identification of emotions [18].

3.1.3. Feature Extraction

The ICare app uses a strong feature extraction methodology to improve the emotion categorization process. Mel-Frequency Cepstral Coefficients (MFCCs) are calculated from audio notifications by the app using Librosa, a robust Python library designed for audio analysis. These coefficients function as important spectral components, reducing the aural qualities of weeping sounds to a format that is easier to handle. Through the extraction of MFCCs from every audio sample, ICare produces characteristic vectors that hold crucial information for the classification of emotions. After that, the machine learning model receives these vectors, which helps it recognize and react appropriately to a range of emotional cues conveyed by the app.

3.1.4. Model Training

As the main component of the machine learning model, a Random Forest classifier is used to aid emotion recognition within the ICare app. Using ensemble learning, this classifier learns a set of rules that together improve the prediction performance of the model by integrating numerous decision trees. The dataset is split into education and check-out sets throughout the training phase to assess the efficacy of the model. The model gains the ability to identify patterns in the MFCC-based full feature vectors and link them to the relevant emotional categories through this procedure. The Random Forest method simulates a collective decision-making process in which each tree, or person, provides their forecast or opinion to the outcome. Because the majority vote serves as the basis for classification, this group decision is solid and trustworthy classification output. Fig. 2 depicts its structure.

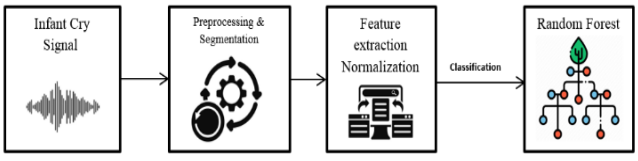

Fig. 2. The process of voice recognition.

3.2 Speech Therapy

The speech treatment method used by the ICare app helps kids from the age of two to five develop smooth, stutter-free speaking. Speaking aloud helps youngsters master speech sounds. Visual clues help them recognize the correct speech targets, and auditory

feedback helps them concentrate on pronouncing words correctly [19].

To accommodate the different developmental phases of children, the therapy is organized into three layers. The first tier introduces animal names and photographs to children aged two to three. For children ages three to four, the second stage concentrates on teaching geometric forms. The third level, which is intended for children ages four to five, is learning verses from the Holy Qur'an by heart. The resemblance between a child's word and the right word was first determined using cosine similarity, but the results were not up to par. The "paraphrase-multilingual-MiniLM-L12-LM" model from Sentence Transformer was then used, which produced word similarity evaluation results that were more tolerable.

3.2.1. Data collection

The approach used to collect the data for this project was to teach the children the names of animals, geometric shapes, and the Holy Quran in addition to improving their speaking skills. The age of the children and the level of age-appropriate complexity of the content were considered when gathering the information.

3.2.2. Audio preprocessing

During the audio preparation phase of the ICare app, data is first collected in a text format that is inappropriate for the purpose for which it was designed. To solve this, Google's text-to-speech capability is used in a conversion process to turn the text data into audio format. After the data is converted to audio, the application waits for a child's input, anticipating that it will come in the form of sound. After that, speech-to-text conversion techniques are used to transform this audio input back into text format so that it may be compared to the current data. The similarity between the child's input and the available data is measured using a comparison metric or similarity measure. Through this approach, the precision and correctness of the child's speech can be evaluated by determining how closely it matches the intended or expected output.

3.2.3. Model Architecture and training

The first model in their investigation employed the Cosine Similarity function shown in Fig. 3, which calculates the similarity between two vectors by evaluating the cosine of their angle. This method has drawbacks even though it works well at capturing semantic similarity using weights or frequencies. It disregards the position or order of characters and may have trouble processing and interpreting lengthy or complicated strings. The researchers used the "Paraphrase-multilingual-MiniLM-L12-LM" pretrained model to overcome these difficulties. As demonstrated in Fig. 4, this model effectively maps phrases and paragraphs to a 384-dimensional vector space, making it suitable for use with 50 languages, including Arabic and English. This allows for efficient storing and retrieval. This model offers a reliable answer for the project's requirements. It consists of a data pooling layer for creating final sentence embeddings and a transformer (a BertModel, to be more precise) with a maximum sequence length of 128 [20]. BERT, which stands for Bidirectional Encoder Representations from Transformers, is a technique for pre-training neural network models to process natural language. Unlike traditional models that process words sequentially, BERT processes words in relation to all other words in a sentence simultaneously, allowing for a better understanding of context and word relationships. In the process illustrated in Fig. 4, BERT is used to compare sentences by representing them as word vectors—arrays of numbers capturing each word's meaning. The cosine similarity between the sentence vectors is then calculated, measuring the similarity between the two vectors on a scale from 1 (perfectly similar) to -1 (perfectly dissimilar). In this context, a high cosine similarity score indicates that the sentences have similar meanings.

The following are the options for the pooling layer: Mean pooling averages the token embeddings; max pooling selects the maximum value across tokens; and CLS token employs the distinct [CLS] token embedding. This is how the final pipeline will appear if the model is configured with the extra preprocessing steps depicted in Fig. 5. Speech therapy addresses various communication issues, including articulation problems, fluency disorders, language delays, and voice disorders. A speech therapy app might feature articulation exercises to help children produce specific sounds correctly by repeating targeted words or phrases, language stimulation activities to enhance vocabulary and grammar skills by naming objects in pictures, and games and puzzles to practice communication skills in an engaging way, such as matching pictures to words. While these apps can be valuable tools for parents and caregivers to supplement professional speech therapy

services at home, they should not replace professional therapy.

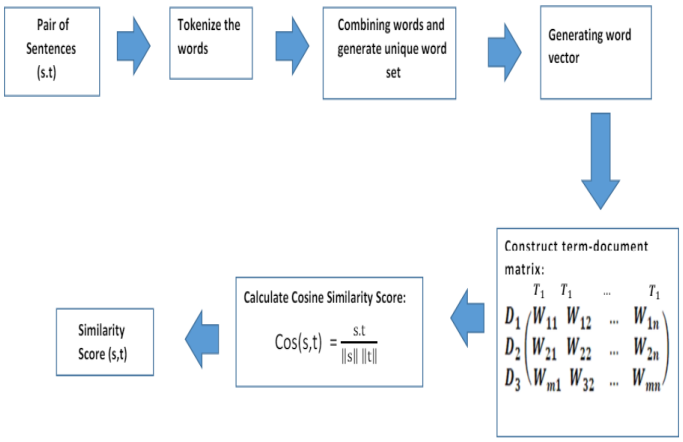

Fig. 3. The steps of Cosine similarity procedure**.**

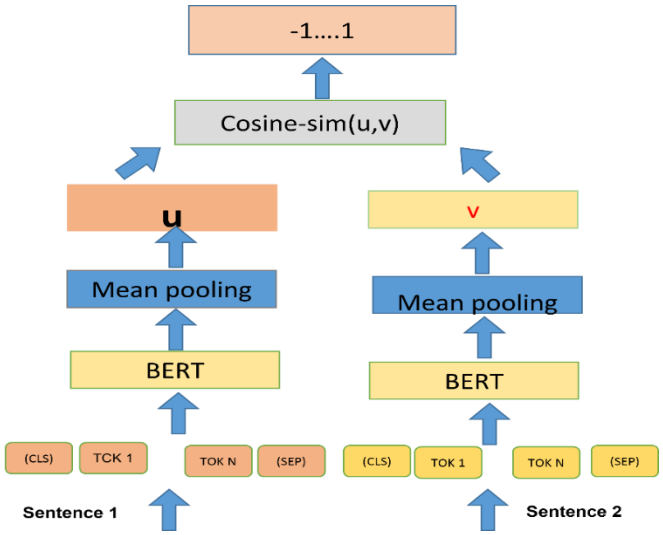

Fig. 4. The structure of the sentence transformer is based on BERT.

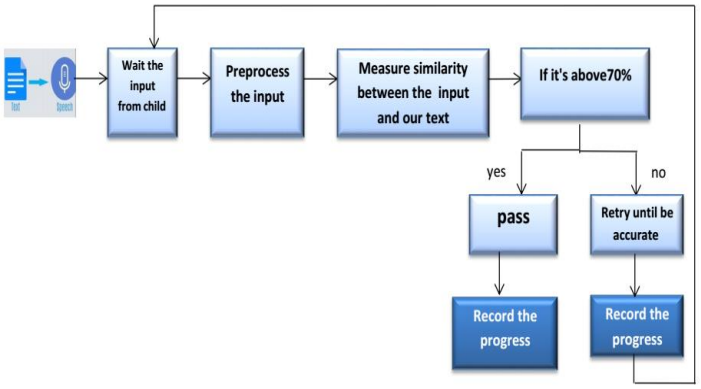

Fig. 5. The process for recording a child's progress in speech therapy.

3.3 Chatbot for healthcare

The ICare chatbot uses artificial intelligence to change childcare by providing personalized support, promoting learning, and fostering healthy development. With ICare, empathy and technology work together harmoniously to support human caregivers rather than replace them. Let's look at how ICare accomplishes this:

- Daily Involvement: ICare engages parents and caregivers through routine questions. Important aspects of a child's existence, such as learning opportunities, emotional well-being, and nutrition, are covered in these questions. By encouraging thoughtful dialogue, ICare keeps parents informed about their children's needs.
- Research-Based Activities: In addition to posing queries, ICare offers activities backed by research. ICare provides practical advice tailored to the child's unique development, whether it's building resilience, introducing new foods, or developing emotional intelligence.
- Deep Insights: To create an exhaustive kid profile, ICare gathers data throughout time. Parents can evaluate their child's growth, identify patterns, and make informed decisions. With the help of this data-driven approach, caregivers may foster a supportive environment.
- Specialized Features: Although ICare's present focus is on parents of children between the ages of 0 and 12, the organization's team is working hard to add features for youngsters with learning needs. Inclusivity is the aim of ICare, whether it is addressing developmental challenges or adjusting to sensory sensitivity.

3.4. A community for sharing and exchanging information

Caregivers can connect with a supportive network for direction, counsel, and information exchange via the ICare program. Members of the community offer practical resources and guidance to help caregivers become more adept at providing care. Additionally, the group offers empathy and emotional support, which reduces loneliness. When caregivers collaborate, they can find innovative solutions to common issues and address challenges. In general, the ICare community strengthens and unites caregivers, giving them the

bravery and fortitude to meet the challenges of caregiving head-on. Our community as depicted in Fig. 6.

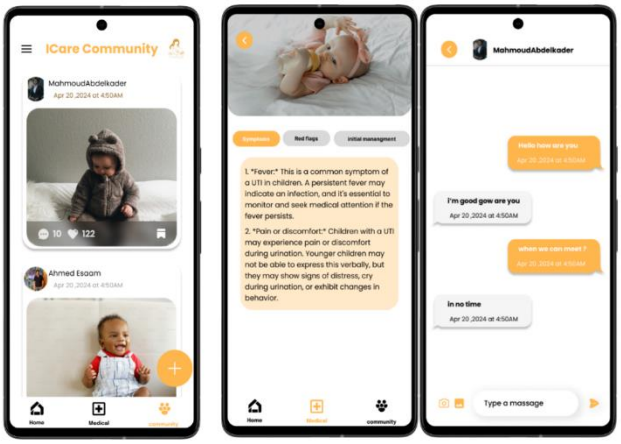

Fig.6. ICare Community.

4. Evaluation Results

The Random Forest model was evaluated using a variety of measures, such as variance, mean accuracy (mAP), and accuracy, throughout the assessment phase. The precision-recall (PR) curve and F1 curve were also used in the evaluation to give a thorough overview of the model's performance. The study made sure to provide an objective appraisal of the model's performance on untested data by dividing the dataset into training and validation sets. The computation of metrics involved comparing the model's predictions with the actual annotations, including accuracy, precision, and recall. While precision represents the fraction of properly predicted positive samples, demonstrating the model's precision, recall indicates the fraction of actual positive samples that the model correctly detected. The evaluation results illustrate the performance of the Random Forest model Fig. 7.

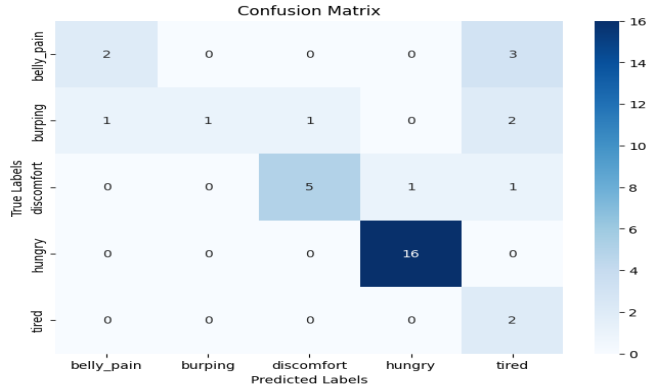

Fig. 7. Confusion Matrix of the proposed method.

According to the categorization report displayed in Fig. 8, This graphic, a bar graph, contrasts three distinct measurements for a total of seven properties. The features belly_pan, burping, pain, hung_gry, tire, Macro_Avg, and Total_Accuracy is indicated on the x-axis. The metrics' values, which range from 0 to 1.2, are displayed on the y-axis.The metrics of Sum of Precision, Sum of Recall, and Sum of F1-score are contrasted in the chart. Different colored bars are used to indicate each metric: orange for Sum of Recall, blue for Sum of Precision, and green for Sum of F1-score. With the highest Sum of Precision, the 'Weighted Avg' function makes the most accurate predictions. Conversely, the 'tire' characteristic has the highest Sum of Recall, indicating that it can accurately detect positive cases with a high degree of accuracy. The 'tire' characteristic also has the greatest Sum of F1-score, suggesting that recall and precision are well-balanced. It is significant to remember that the Sum of F1-score, Sum of Precision, and Sum of Recall change for every feature. Across all three criteria, no particular trait is the most prominent. This diagram can be used to analyze performance or pinpoint areas where a system or procedure needs to be improved. And our total accuracy overall is 74%.

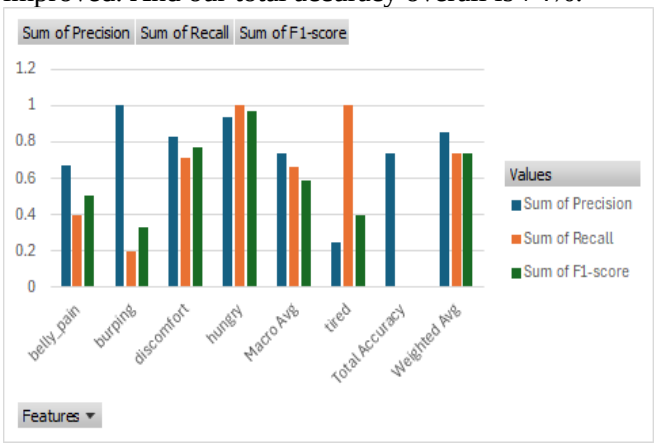

Fig.8. The results of the proposed RF methodology.

Fig. 9 presents a comparison of the proposed model, paraphrase-multilingual-MiniLM-L12-v2, with some of the most recent techniques in the field. The purpose of the supplied bar chart is to compare the typical performance of various paraphrasing models. Four models are shown on the X-axis of the chart: a proposed paraphrase model, distiluse-base-multilingual-cased-v1, distiluse-base-multilingual-cased-v2, and paraphrase-MiniLM-L3-v2. With a range of 0 to 80, the Y-axis most likely reflects a performance statistic like accuracy or F1 score. Each of the four models has an average performance indicated by blue bars, with scores ranging from the mid-50s to the 70s. Notably, the highest

average performance is exhibited by the proposed paraphrase model. A legend on the graphic states that orange bars would indicate "Model Size" and blue bars would show "Avg. Performance," although the orange bars are absent from the illustration. This absence renders the chart incomplete, as does the absence of a title. If model size is relevant, the chart should contain orange bars for completeness and clarity. It should also have a title that clearly states the chart's goal. Overall, the graph shows that, when it comes to average performance using the specified metric, the suggested model performs better than the other models.

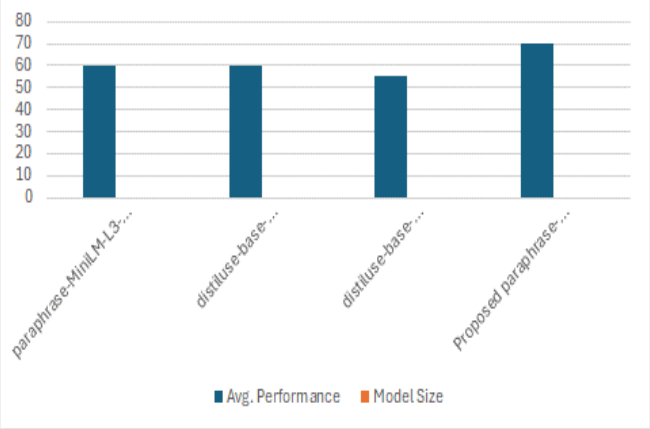

Fig. 9. Speech therapy performance.

Using Gemini API: AI Dart SDK: With the Google AI Dart SDK, developers can create AI-powered features and applications by leveraging Google's cutting-edge generative AI models, such as Gemini.

For prototyping purposes only, it is advised to use the Google AI SDK for Dart (Flutter) to contact the Google AI Gemini API directly from within your app. We highly advise using the SDK to use the Google AI Gemini API solely server-side if you intend to enable billing in order to protect your API key. If you integrate your API key directly into your mobile or web application or retrieve it remotely during runtime, you run the risk of making it known to malicious parties. And Fig. 10 displays our Chatbot for Healthcare.

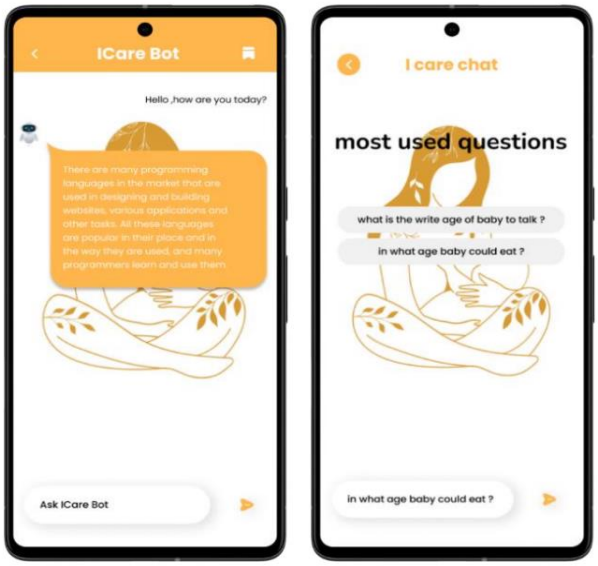

Fig. 10. ICare Chatbot.

The image displays the architecture of a MongoDB deployment that is sharded. The key element of this setup is the MongoDB query router, or mongos, which receives requests from clients. The mongos instance acts as a middleman by utilizing the sharding key to select which shard to direct client queries to. MongoDB's load distribution allows it to handle large datasets and high-performance operations because each shard is a subset of the database. The architecture displayed consists of three shards, each consisting of a parent node and two subsidiary nodes. Each shard consists of a primary node that controls write operations and secondary nodes that replicate data, provide redundancy, aid in failover, and ease read operations. Additionally, the system features replica sets of three nodes called config servers, which are typically used to store information and configuration settings for the sharded cluster. The metadata includes information on how the data is split up among the shards. Client requests are routed to the appropriate shard by the mongos instance. Following the request processing, the data is written to the selected shard's primary node, which subsequently replicates it to the shard's subsidiary nodes. The mongos instance ultimately provides the response to the client. Massive volumes of data and high traffic can be handled with ease by MongoDB because of its high availability and horizontal scaling architecture. as depicted in Fig. 11.

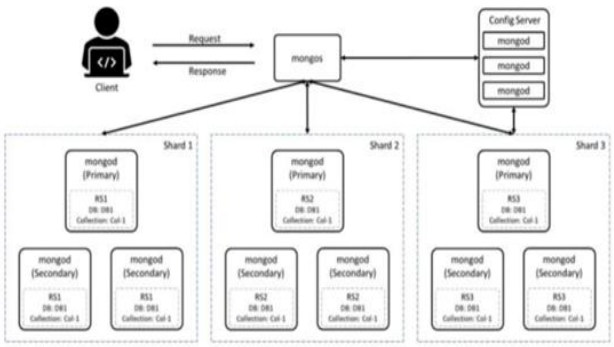

Fig.11. Sharded MongoDB Architecture Diagram.

5. Conclusion and Future Work

The creation of the ICare smartphone application, which gives parents a vital tool to efficiently nurture and assist their young children, marks a ground-breaking milestone in infant care. With its many features, ICare provides parents with important tracking and guidance so they can make well-informed decisions regarding their children's wellbeing. One of ICare's most amazing features is its application of artificial intelligence to interpret baby cries, assisting parents in understanding the causes of their child's discomfort and promoting improved responsiveness and communication. Parents can learn more about many facets of infant development via ICare's conversation module, which will help them better understand and bond with their child. By identifying possible allergens and providing avoidance techniques, ICare also addresses important topics like allergy management, guaranteeing a safer environment for kids. The app's Band-Aid teaching tool gives parents the know-how to deal with common ailments and mishaps, increasing their confidence in handling a variety of circumstances. The app's predictive features, along with its communication feature, allergy control, and instructional elements, make it a useful and essential resource for parents. The ICare's AI-powered features represent a major advancement in the use of AI to enhance the human experience in childcare. The classification report shows a 74% model accuracy, with excellent precision and recall for certain classes such as "hungry," and highlights areas that require more development, including the "tired" class. Future work will concentrate on improving the model's accuracy in these domains and developing the application's functionalities to accommodate a wider age range and a greater variety of data. The user interface and overall functionality will be improved based on ongoing user feedback, guaranteeing that ICare will continue to be an essential resource for parents throughout the world.

References:

[1] Sirbu, A.: The significance of language as a tool of communication. Sci. Bull. Mircea Cel Batran Nav. Acad. 18, 405 (2015).

[2] Bird, J., Bishop, D.V.M., Freeman, N.H.: Phonological Awareness and Literacy Development in Children with Expressive Phonological Impairments. J. Speech Lang. Hear. Res. 38, 446–462 (1995). https://doi.org/10.1044/jshr.3802.446.

[3] Conti-Ramsden, G., Botting, N.: Social Difficulties and Victimization in Children with SLI at 11 Years of Age. J. Speech Lang. Hear. Res. 47, 145–161 (2004). https://doi.org/10.1044/1092-4388(2004/013).

[4] Simon, P., Rosenbaum, S.: Speech and language disorders in children: Implications for the social security administration's supplemental security income program. (2016).

[5] Hulme, C., Snowling, M.J., West, G., Lervåg, A., Melby-Lervåg, M.: Children's Language Skills Can Be Improved: Lessons from Psychological Science for Educational Policy. Curr. Dir. Psychol. Sci. 29, 372–377 (2020). https://doi.org/10.1177/0963721420923684.

[6] Chang, C.-Y., Tsai, L.-Y.: A CNN-Based Method for Infant Cry Detection and Recognition. In: Barolli, L., Takizawa, M., Xhafa, F., and Enokido, T. (eds.) Web, Artificial Intelligence and Network Applications. pp. 786–792. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-15035-8_76.

[7] Nadia Maghfira, T., Basaruddin, T., Krisnadhi, A.: Infant cry classification using CNN – RNN. J. Phys. Conf. Ser. 1528, 012019 (2020). https://doi.org/10.1088/1742-6596/1528/1/012019.

[8] Ji, C., Basodi, S., Xiao, X., Pan, Y.: Infant Sound Classification on Multi-stage CNNs with Hybrid Features and Prior Knowledge. In: Xu, R., De, W., Zhong, W., Tian, L., Bai, Y., and Zhang, L.-J. (eds.) Artificial Intelligence and Mobile Services – AIMS 2020. pp. 3–16. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-59605-7_1.

[9] Turan, M.A.T., Erzin, E.: Monitoring Infant's Emotional Cry in Domestic Environments Using the Capsule Network Architecture. In: Interspeech 2018. pp. 132–136. ISCA (2018). https://doi.org/10.21437/Interspeech.2018-2187.

[10] Foote, K.D.: A Brief History of Natural Language Processing, https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/, last accessed 2024/06/08.

[11] Oh, B.-D., Lee, Y.-K., Kim, J.-D., Park, C.-Y., Kim, Y.-S.: Deep Learning-Based End-to-End Language Development Screening for Children Using Linguistic Knowledge. Appl. Sci. 12, 4651 (2022). https://doi.org/10.3390/app12094651.

[12] Heilmann, J., Miller, J.F.: Systematic Analysis of Language Transcripts Solutions: A Tutorial. Perspect. ASHA Spec. Interest Groups. 8, 1–18 (2023). https://doi.org/10.1044/2022_PERSP-22-00148.

[13] Reimers, N., Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, http://arxiv.org/abs/1908.10084, (2019). https://doi.org/10.48550/arXiv.1908.10084.

[14] Behbehani, Farah, Alysse J. Kowalski, Helina Selam, Eileen Dombrowski, and Maureen M. Black. "Childcare centre attendance and health, growth, and development among children aged 0–3 years in low-and middle-income countries: A systematic review." Journal of Global Health 14 (2024).

[15] Sarhan, S., Nasr, A.A., Shams, M.Y.: Multipose Face Recognition-Based Combined Adaptive Deep Learning Vector Quantization. Comput. Intell. Neurosci. 2020, (2020).

[16] Ji, C., Mudiyanselage, T.B., Gao, Y., Pan, Y.: A review of infant cry analysis and classification. EURASIP J. Audio Speech Music Process. 2021, 8 (2021). https://doi.org/10.1186/s13636-021-00197-5.

[17] Veres, G.: gveres/donateacry-corpus, https://github.com/gveres/donateacry-corpus, (2024).

[18] Shams, M.Y., Hassanien, A.E., Tang, M.: Deep Belief Neural Networks for Eye Localization Based Speeded up Robust Features and Local Binary Pattern. In: Shi, X., Bohács, G., Ma, Y., Gong, D., and Shang, X. (eds.) LISS 2021. pp. 415–430. Springer Nature, Singapore (2022). https://doi.org/10.1007/978-981-16-8656-6_38.

[19] Keith, M., Wiliamson, S., Sotto, C.: The Importance of Repetition and Practice of Speech Sounds in Elementary School-Aged Children. Undergrad. Sch. Showc. (2019).

[20] Hassan, E., Abd El-Hafeez, T., Shams, M.Y.: Optimizing classification of diseases through language model analysis of symptoms. Sci. Rep. 14, 1507 (2024). https://doi.org/10.1038/s41598-024-51615-5.