International Journal of Current Bio-Medical Engineering (IJCBE)
Volume 1, Issue 2, November – December (2023)

RESEARCH ARTICLE

CLJFS: DEEP LEARNING BASED CLASSIFICATION OF JAUNDICED INFANT SIGNALS

J. Jency [1, *], J. Anto Germin Sweeta [2], M. Annies Stelina [3]

[1] Research Scholar, Department of Computer Science, S. T. Hindu College, Nagercoil, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, India

[2] Research Scholar, Department of Computer Science & Research Centre, S.T. Hindu College, Nagercoil – 629002, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli – 627012, Tamil Nadu, India.

[3] Research Scholar, Department of Computer Science & Research Centre, S.T. Hindu College, Nagercoil – 629002, Affiliated to Manonmaniam Sundaranar University, Tirunelveli – 627012, Tamil Nadu, India.

Corresponding e-mail: jencysekar1995@gmail.com

Abstract – Premature newborns cry to let their parents and others know what they need. It is essential to remember that premature screamers may scream for a variety of causes. Based on weeping, parents can identify a baby's emotional and physical changes and needs. However, quite challenging to pinpoint the requirement for the incubated newborns that have jaundice. To overcome these challenges, a novel deep learning-based CLJFS model is proposed for Cry-based Jaundiced Infant Signal (CLJFS) classification model. The crying is a newborn's primary communication, understanding its acoustic features can provide critical insights into the infant's condition. The proposed CLJFS model employs a multi-step process beginning with signal pre-processing using Stationary Wavelet Transform (SWT) for noise reduction and feature enhancement. Linear Prediction Coefficients (LPC) are extracted, followed by feature selection using the Least Absolute Shrinkage and Selection Operator (LASSO). A Spiking Neural Network (SNN) then categorizes the cries into three classes: hunger, fear, and discomfort. The effectiveness of the proposed CLJFS was evaluate using F1 score, accuracy, precision, recall, and specificity. The proposed CLJFS model achieved a classification accuracy 98.9%. The proposed model enhanced the total accuracy by 2.24%, 4.03%, 10.1%, and 7.03%, respectively.

Keywords *– Infant cry signal, deep learning, Stationary Wavelet Transform, Linear Prediction Coefficients, Spiking Neural Network.*

1. INTRODUCTION

Crying is the first form of communication for newborns, serving as their primary means to express physiological needs, emotions, or discomforts. This complex act requires coordination between the brain, respiratory system, muscles, and vocal cords. [1] For new parents and caregivers, interpreting these cries can be challenging, as the reasons behind the cries may range from hunger, fatigue, or discomfort to medical conditions such as jaundice [2]. Advances in understanding the acoustic features of infant cries have revealed that they carry linguistic signals unique to specific needs, but distinguishing them often remains subjective, especially in clinical settings [3, 4].

Neonatal jaundice, a common condition affecting over 80% of newborns, is caused by elevated bilirubin levels in the blood [5, 6]. This condition typically manifests within the first five days after birth and is marked by yellowing of the skin and tissues. It results from an immature hepatic system and heightened hemoglobin breakdown, particularly in preterm infants [7]. Factors such as insufficient milk intake and genetic predispositions exacerbate the delayed clearance of bilirubin, necessitating timely diagnosis and management [8]. Screening tools like Transcutaneous Bilirubin (TcB) or Total Serum Bilirubin (TSB) are essential for early detection and prevention of complications such as neurotoxicity.

Understanding neonatal jaundice is critical, especially for preterm infants or those with specific risk factors such as hemolysis, sepsis, or glucose-6-phosphate dehydrogenase (G6PD) deficiency [9]. Addressing the condition requires identifying underlying causes and implementing strategies like phototherapy or exchange transfusion when bilirubin levels exceed established thresholds. Effective management not only minimizes the risk of neurotoxicity but also ensures the infant's healthy development [10]. For healthcare providers and parents, combining medical interventions with breastfeeding support and vigilance during the critical postnatal period is essential.

Newborns communicate primarily through crying, making it challenging for caregivers to identify the exact cause of distress, particularly in medical conditions like neonatal jaundice [11]. Jaundice, caused by elevated bilirubin levels, is common in infants and requires timely detection and treatment to prevent complications like neurotoxicity. [12] The subjective interpretation of infant cries highlights the need for more reliable, objective tools for diagnosing and managing neonatal health issues effectively. To overcome this problem, a novel CLJFS model is proposed for classifying the preterm infant cry signals. The major research contributions are mentioned below:

- The main goal of this research CLJFS model for classifying preterm babies' cry signals in incubators.
- In Stationary Wavelet Transform (SWT), an infant cry signal is pre-processed by dissecting the signal into several frequency components.
- Using deep learning-based ShuffleNet, the Linear Prediction Coefficients (LPCs) characteristics from the pre-processed signals are extracted, and the retrieved features are then chosen using the Least Absolute Shrinkage and Selection Operator (LASSO).
- The Spiking Neural Network (SNN) will categorise the signal into three categories: hunger, fear, and discomfort using the output from the feature selection phase and the bilirubin monitoring system.
- Accuracy, precision, specificity, recall, and F1 score are some of the measures used to evaluate the performance of the proposed.

The remainder of the work is organized accordingly. The literature review is summarized in Section 2. Section 3 discusses about the proposed model and Section 4 focuses at the performance of the proposed approach and compares it to other methods. Section 5 provides a final explanation of the conclusion and future scope.

2. LITERATURE SURVEY

In this section, the state-of-the-art in the domains that are pertinent to the work that is being presented, including data, deep learning, machine learning methods, and related studies. Related studies are listed in the following paragraphs, with an emphasis on those that make use of categorize infant baby cries.

In 2019 Maghfira, T. N., et al., [13] investigated the challenge of categorizing five different sorts of emotions or demands communicated by newborn cries, including signs of hunger, tiredness, stomachache, discomfort, and the need to burp. The suggested CNN-RNN model performs better than the prior approach by an average classification accuracy of up to 94.97%, according to analysis of the Dunstan Baby Language dataset.

In 2019 Severini, M., et al., [14] examined by the cry detection issue in medical settings such as Neonatal Intensive Care Units (NICUs). The article discusses several DNN-based single- and multi-channel solutions. After processing signals using a post-filter, a beam former, and a single-channel DNN, the assessment showed that training with actual data enables the attainment of the highest possible overall performance, with a PRC-AUC of 87.28%.

In 2022 Joshi, V. R., et al., [15] devised for effective answers to the conundrum of figuring out what causes a baby to scream. The integration of multiple sophisticated boosting algorithms at different levels was its main benefit. The suggested model finally showed excellent performance with an average classification accuracy of as much as 93.7% after several comparisons.

In 2022 Liang, Y. C., et al., [16] suggested using deep learning (DL) protocols in this study to identify infants' physiological requirements, including the need for food and drink, the need to change diapers, emotional needs like the need to be held or touched, and discomfort from treatments like shots. CNN outperformed LSTM and ANN across practically all metrics, achieving up to 60% accuracy in identifying newborns' unique demands.

In 2022, Lahmiri, S., et al., [17] designed and validated several deep learning methods to enhance the diagnosis of baby cry recordings which is the aim of the current work. The number of neurons in the hidden layers and the number of convolutional layers in CNN and DFFNN are adjusted in tandem with one another. In comparison with related works in the literature, Cepstrum analysis-based coefficients taught DL systems are powerful instruments that may be applied to precise diagnosis of newborn cry records to separate healthy and abnormal signals.

In 2022 Cha, J., & Bae, G. [18] proposed deep learning to categorize videos of crying babies. The approach uses autoencoder, deep residual network, and concatenate layer classification models as well as a variety of audio methods for extracting features (Short-Time Fourier Transform and Mel Frequency Cepstral Coefficient). The experiment's findings demonstrate that, in comparison to previous machine learning-based models, the suggested model achieves excellent accuracy in reading newborn cries.

In 2022 Ashwini, K., & Vincent, P. D. R. [19] uses a Deep Convolutional Neural Network to transform audio inputs into spectrogram images in order to create a prediction model for a newborn cry categorization system. A deep neural network-based newborn cry classification system can classify sleeping cries with an accuracy of up to 95%. SGDM optimization with Convolutional neural network achieves greater prediction accuracy, according to the results.

In the literature review, above existing techniques had developed using various DL and ML approaches to infant cry signals. However, existing methods for infant cry signal classification often suffer from limited accuracy due to inadequate feature extraction and preprocessing techniques. Additionally, many models fail to effectively handle the variability and noise inherent in real-world cry signal datasets. In this research, CLJFS method was proposed for classifying the infant baby cries.

3. PROPOSED METHODOLOGY

This section proposes a unique method for cry signal categorization of preterm infants in incubator called Neonatal Advanced Multisource Experiment with Cry Signal (CLJFS). A baby's cry signal is pre-processed, the features are retrieved using Linear Prediction Coefficients (LPCs), and then LASSO selects the features. A bilirubin monitoring system that tracks and forecasts bilirubin levels is part of the suggested system. The Spiking Neural Network (SNN) will categorize the signal into three categories: hunger, fear, and discomfort using the output from the feature selection stage and the bilirubin monitoring system. Finding these signals aids in pinpointing the origin of newborn crying.

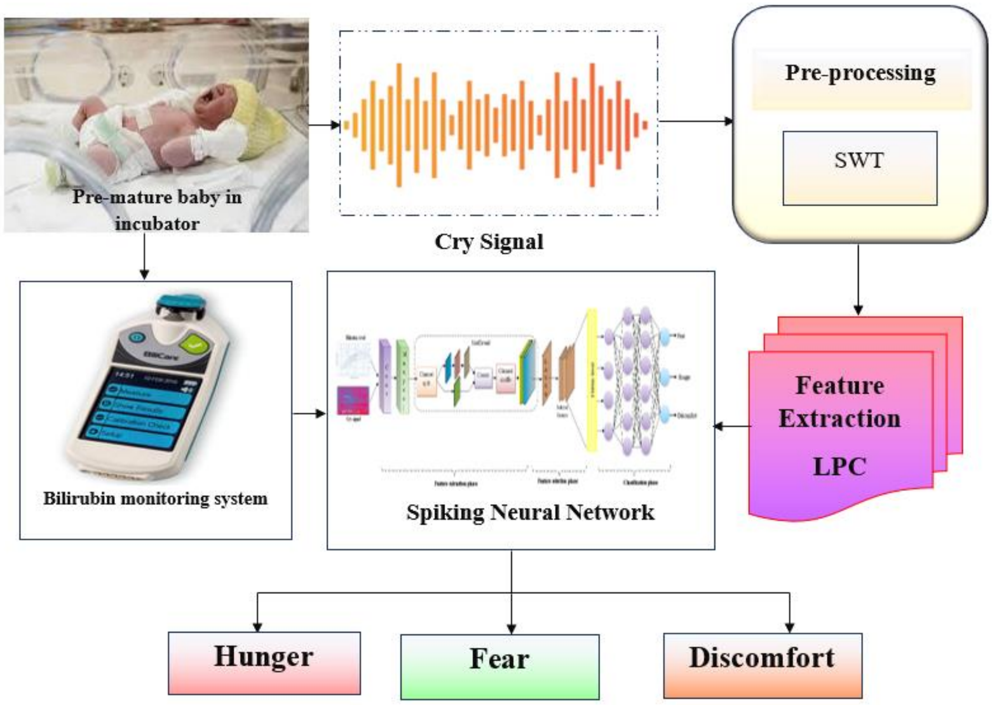

Figure 1. The Overall Workflow of the Proposed Methodology

3.1 Dataset Description

The NICU data from adjacent hospitals and neighboring colonies are used to create the database [11]. Both at home and at the clinic, the baby's activity levels are constantly tracked. Both parents and hospital employees have given their approval for the data collecting. The database contains 27 sound signal streams, 12 from different neonates and the rest from babies who are similar. For both men and women, distinct voices may be heard in this audio.

3.2 Pre-processing

Baby cries are recorded and labeled by caregivers or medical professionals in homes or hospitals using digital recorders, but the databases are small due to the sensitive data collection process. Pre-processing involves audio segmentation and denoising to remove background noise like conversation or footsteps, ensuring accurate analysis. Stationary Wavelet Transform (SWT) is used to decompose cry signals into frequency subbands, preserving time invariance and improving feature extraction for classification.

The input medical image (I) index set displays the pixels in the xth column and yth row, which is thought of as $2D[x,y], I[x,y]$. To produce the vertical coefficient (LH), approximation coefficients (LL), horizontal coefficient (HL), and diagonal coefficient (HH), SWT performs first level 2D-SWT on the picture. Using the 2DSWT method, two wavelet subbands were recovered from the medical picture, one for each subband coefficient of the wavelet transform. The following is a representation of the 2DSWT's approximate and detailed coefficients:

$$\tilde{c}_{i+1,j,n} = \sum_{U=-\infty}^{\infty} h(u)h(u)\tilde{c}_{i,j+2^i,n+2^iV}, \quad (1)$$

$$\tilde{d}_{1,i+1,j,n} = \sum_{U=-\infty}^{\infty} h(u)h(v)\tilde{d}_{1,i,j+2^i,n+2^iV}, \quad (2)$$

$$\tilde{d}_{2,i+1,j,n} = \sum_{U=-\infty}^{\infty} h(u)h(v)\tilde{d}_{2,i,j+2^i,n+2^iV}, \quad (3)$$

$$\tilde{d}_{3,i+1,j,n} = \sum_{U=-\infty}^{\infty} h(u)h(v)\tilde{d}_{3,i,j+2^i,n+2^iV}, \quad (4)$$

Where, respectively, $c_{i,j}$ and $d_{i,j}$ stand for approximate and detailed coefficients. Following 2DSWT decomposition, the concatenation of the four subbands is always the same size as the original input. 2D inverse SWT (2DISWT) may be traced back to 2DSWT by reversing the techniques.

3.3 Feature Extraction

Infant cries contain vocalizations, silences, and interruptions, reflecting physiological and prosodic information. Feature extraction, a key step in machine learning, involves deriving discriminative features from cry signals in time, frequency, or cepstral domains. Time-domain features like amplitude and energy are simple but prone to noise, while frequency-domain features, including MFCCs and LPCs, better capture signal variations and cyclic patterns. Cry signals are segmented into overlapping frames using a Hamming window to ensure quasi-stationarity, enabling short-term analysis. Autocorrelation and LPC coefficients are calculated for each frame, supporting effective cry signal characterization. Combining acoustic, prosodic, and image-based features, spectrograms enhance discrimination and classification. Advanced techniques like ShuffleNet, with depth-wise separable convolutions, reduce computational costs while maintaining high performance, serving as an efficient encoder for transforming cry signal features. This multi-dimensional approach improves cry signal analysis and classification, especially for medical applications.

$$fps = 2 \times d \times h \times w \times (C_{in} \times k^3 + 1) * C_{out} \quad (5)$$

where the output feature maps' height, width, and depth are denoted by h, w and d, respectively; the input and output streams are denoted by C_{in} and C_{out}; and the kernel size is denoted by k. ShuffleNet uses a depth-separable convolutional layer to modify the key feature channels.

$$F_c = l_c + \sum_{k=1}^{c-1} \left[{k}/{c} \right] \times F_c l_{c-k} \quad (6)$$

Here, the cepstrum coefficient is defined as F_c, and l_c stands for the linear prediction coefficient. LPC are not easily influenced by artefact. Compared to LPC features, LPC features provide reduced mistake rates. Due to theoretical restrictions, the higher-level variances of cepstral factors were higher as we traverse through lower to higher level cepstral factors.

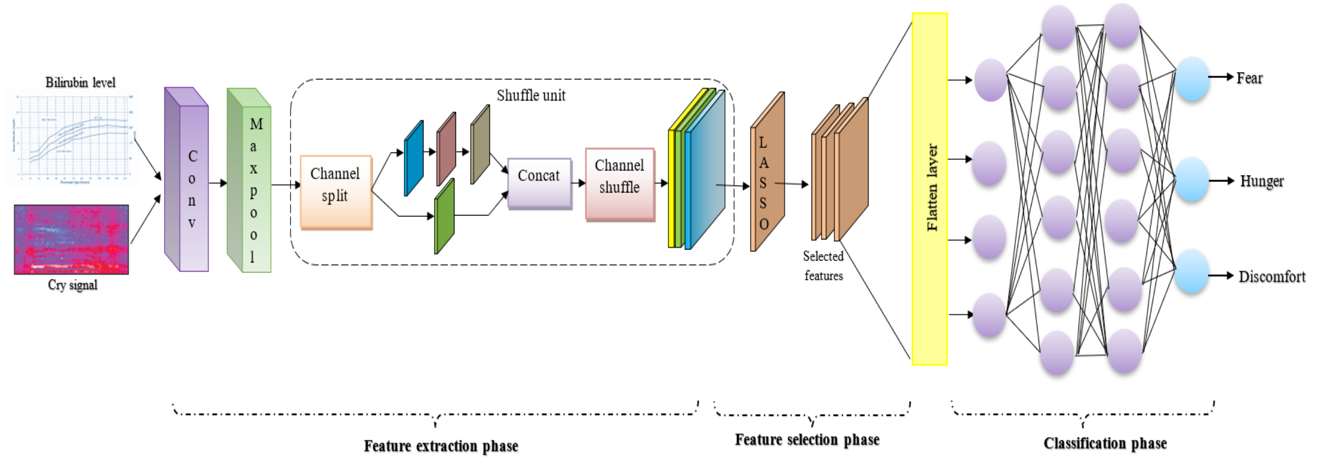

Figure 2. The architecture of Shuffled Spiking Neural Network

The LPC determinations are notorious for being extremely vulnerable to quantization noise. High-pitch speech signals are difficult to distinguish from source-filters employing cepstral evaluation in the frequency domain. Higher-order cepstral factors are impacted by noise, but the lower-order cepstral coefficients are responsive to spectral slope. The feature selection phase uses the retrieved LPC features to choose the most pertinent characteristics for categorizing the cry signal.

3.4 Feature Selection

LASSO (Least Absolute Shrinkage and Selection Operator) is a machine learning technique used for feature selection and linear regression. By adding an L1 regularization term to the cost function, LASSO reduces the coefficients of less important features to zero, effectively selecting the most relevant ones. This reduces model complexity, prevents overfitting, and ensures interpretability. The technique is particularly effective for high-dimensional data, as it produces sparse solutions by including only essential features. However, it performs better when the number of features is lower than the sample size. By minimizing redundant features and optimizing the Ordinary Least Squares (OLS) loss, LASSO enhances model generalization and efficiency. The LASSO algorithm then learns the coefficients of sparse regression as

$$L_0 = argmin_\beta \; ||x - y \times L||^2 + \gamma ||\mathrm{L}|| \; sel_f \quad (7)$$

Where the response vector is equal to $x_1, x_2 \ldots \ldots x_n$, the feature matrix is equal to $y = y_1, y_2 \ldots \ldots y_n$, and the trade-off parameter is equal to $\gamma = 0.005$ for comparing L_0's sparsity and fitting efficiency. The effect of sparse coefficient valuation is controlled by the regularisation constraint. The three types of infant cries are fear, hunger, and discomfort. After that, the best feature subset is flattened and used as input for the classification process.

3.5 Classification

Spiking Neural Network may achieve remarkable accuracy with few synapses, Spiking Convolutional Neural Network architecture has drawn a lot of interest recently. Compared to CNN, SNN uses a substantial amount less energy. Average pooling, convolution, and other standard network layers are supported by SNN. For the SNN operations, the input cry signal must first be converted into spikes. The suggested SNN consists of a pooling layer, a fully-connected layer, and a hierarchy of convolutional layers. SNNs depend on discrete occurrences occurring at specific moments in time rather than continuous values. Some neurons in the SNN model take longer to activate, which increases their quality value. Neuron threshold is exceeded, it activates, influencing neighboring neurons, while failure to exceed the threshold reduces activity temporarily. Activation is signaled by acute spikes, with timing determined by pulse intervals. SNNs rely on coders to encode signals into spike patterns and decoders to translate these patterns into numerical outputs. Neuron activation depends on crossing the threshold, with outputs governed by pulse frequency and range in the coding scheme. Decoding converts spike codes into numerical values for processing.

Equation (8) is used to calculate the pulse value of the SNN via the Leakage Integral and Fire algorithm (LIF).

$$I_u(t) - \frac{V_u(t)}{R_u} = C_u \frac{\partial V_u(t)}{\partial t} \quad (8)$$

Where I_u stands for "current unit," V_u, for "volt unit," C_u for "capacitor unit," R_u for "resistor unit," and t for "time unit." An SNN-based classifier is used in supervised learning models. Each learning instance corresponds to one output neuron in the SNN, whose neurons are linked. In the rank-order learning process, the synaptic weight between neurons of i and j is calculated,

$$wt_{i,j}(t) = mod^{order(i,j)} \tag{9}$$

where $order(i,j)$ denotes the initial order of the received spike and $wt_{i,j}$ denotes the associated synapse between neurons i and j. The spike-based synaptic plasticity learning rule modifies the synaptic weight using a drift parameter. As a result, the synaptic weights are altered while taking the incidence of future spikes over time, or $Sp_i(t)$, naturally into account. In order to update the weights, equation (10),

$$wt_{i,j}(t) = \begin{cases} wt_{i,j}(t-1) + drift, & Sp_j(t) = 1 \\ wt_{i,j}(t-1) + drift, & Sp_j(t) = 0 \end{cases} \tag{10}$$

where $Sp_j(t)$ is the time t's spike. If neuron i spikes at time t, the synaptic weight is potentiated as opposed to reduced if neuron i spikes after the presynaptic neuron fires. Benefits of the SNN include the ability to categorize the spatiotemporal cry signal into categories like hunger, fear, and discomfort.

4. RESULTS AND DISCUSSIONS

The proposed CLJFS approach and its efficiency based on LPC characteristics have been analyzed through Matlab-2019b. The obtained cry signals collected through [11], and the corresponding wave patterns are preprocessed and converted as frames to enhance subsequent processes. The spectrogram signal having prosodic characteristics obtained from actual wavelet patterns have been categorized using CLJFS. It is possible to compute the assessment of the test cases using metrics like F1-score, recall, exactness, specificity, and correctness.

4.1 Performance analysis

The proposed CLJF Scan be evaluated based on accuracy, specificity, precision, recall, and F1 score. The effectiveness of the suggested network for classifying various cry signal kinds is shown in table 1 and is visually depicted in Figure 3.

Table 1. Performance assessment of the CLJFS model

Classes	Precision	Specificity	Recall	Accuracy	F1 score
Hunger	97.02	96.24	96.25	99.14	98.24
Fear	96.12	95.71	95.05	98.25	97.68
Discomfort	98.27	97.04	97.15	99.51	98.24

The effectiveness of the suggested CLJFS for categorizing various cry signals, such as hunger, fear, and discomfort, is shown in the table 1. The specificity, precision, recall, f1 score, and accuracy of the CLJFS model were used to determine its efficacy. The suggested model has a 98.9% total accuracy. The total specificity, precision, recall, and F1 score for the proposed CLJF Sare 96.33%, 97.13%, 96.15%, and 98.05%, respectively

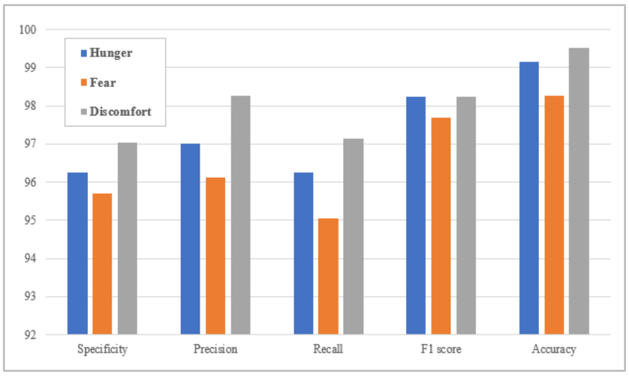

Figure 3. Graphical representations for different emotions of the premature baby

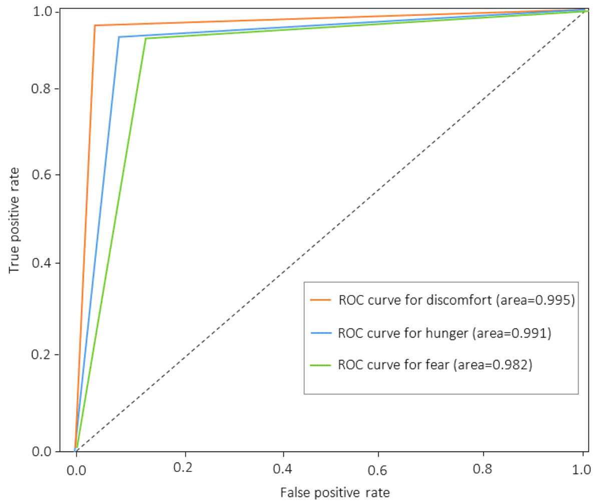

Figure 4. ROC curve of the proposed CLJFS model

Figure.4 shows the ROC produced for several classes of preterm newborn cries using the acquired dataset that obtains a better AUC. With parameters like TPR on the y-axis and FPR on the x-axis, the suggested CLJFS obtained AUC of 0.991 for hunger, 0.982 for fear, and 0.995 for pain.

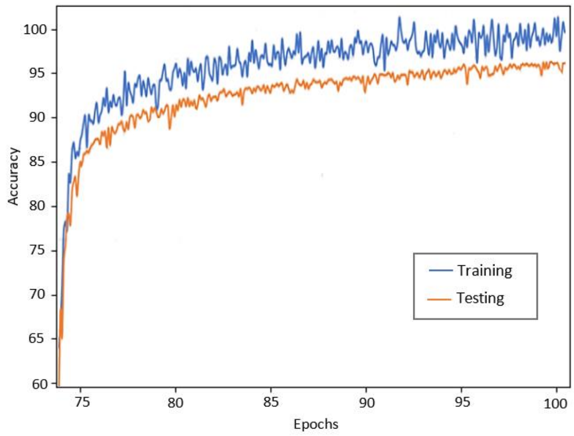

Figure 5. Accuracy curve of the proposed CLJFS model

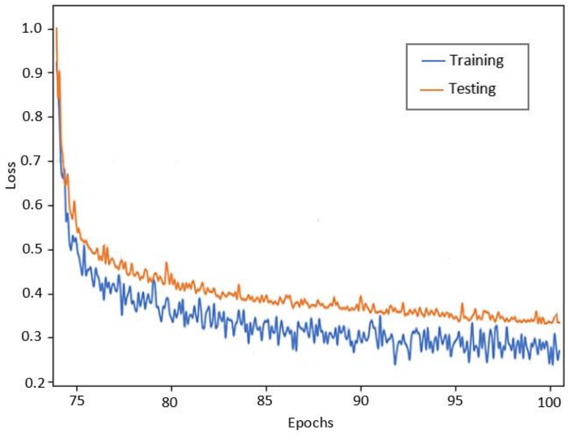

Figure 6. Loss curve of the proposed CLJFS model

The spectrum of precision on the vertical axis as well as the number of epochs on the horizontal axis may be used to visualize the accuracy curve shown in Figure 5. As the number of epochs increases, so does the accuracy of the CLJFS. Figure 6 illustrate the epoch and loss range, which demonstrates that the loss of the CLJFS is decreased as the epoch rises. The suggested CLJFS achieves great accuracy for recognizing the various cry signal types. This study is currently determining how many epochs are required to get a high degree of accuracy. From the fallouts, the minimum error rate was achieved at 98.9% testing accuracy, leading to the classification accuracy of CLJFS being attained after 100 training epochs.

4.2. *Comparative analysis*

The effectiveness of each Deep Learning network was evaluated in order to confirm that the CLJFS achieves high accuracy in its outputs. The proposed CLJFS was compared with deep learning classifiers including AlexNet, GoogleNet, and MobileNet in a contrast analysis. The obtained accuracy value using the proposed SNN is 98.9%, that has been higher than the traditional DL models for the feature extraction process. Performance estimation was done using several measures, including precision, recall, specificity, f1 score, and accuracy of each DL network.

Table 2. Comparison among classic deep neural networks for feature extraction

Networks	Parameters	Flops	Complexity	Accuracy
AlexNet	56980k	722 M	720	90.2%
GoogleNet	7544k	525 M	140	98.01%
MobileNet	81.45k	568 M	170	96.68%
ShuffleNet	17.3k	38M	38	98.96%

As shown in table 2, traditional networks like AlexNet, GoogleNet, and MobileNet are more complex since they need numerous parameters to attain high accuracy. ShuffleNet, on the other hand, employs a smaller number of parameters, which reduces complexity while retaining high accuracy levels of 98.9%. When compared to conventional networks, complexity has been reduced by a factor of four. Table 3 displays a comparison analysis of earlier deep learning networks for classification.

Table 3. Comparative Analysis among Traditional models for classification

Networks	Precision	Specificity	Recall	F1 score	Accuracy
GoogleNet	95.6	95.4	96.8	96.4	95.2
MobileNet	95.0	97.4	97.5	98.5	97.8
AlexNet	86.7	88.1	85.6	87.1	90.2
SNN	97.1	96.3	96.1	98.0	98.9

Table 3 analyses many DL approaches based on certain performance criteria, with the goal of reaching the widest accuracy range in the categorization phase of premature newborn scream. Additionally, compared to SNN, conventional networks do not reach great accuracy. Compared to AlexNet and GoogleNet, the SNN improves the total accuracy range by 8.7%, 3.7%, and 1.1%.

Table 4. Accuracy comparison of state-of-art models and Proposed model

Authors	Method	Accuracy
Ji, C., et al (2019)	DL based merged feature matrix	96.74%.
Maghfira., et al (2020)	CNN-RNN	94.97%
Ashwini., et al (2021)	SVM-RBF	88.89%
Anjali., et al (2022)	Finetuned VGG16	92.0%
Proposed	CLJFS	98.96%

Table 4, the practical length of evaluation images from the obtained dataset in the testing phase was tallied to evaluate the efficacy of various methods. Modern models were compared against one another using certain performance metrics while maintaining the necessary classification accuracy. In comparison to CNN-RNN, SVM-RBF, CNN-RNN and Finetuned VGG16, the CLJFS improves overall accuracy by 2.24%, 4.03%, 10.1%, and 7.03%, respectively. Therefore, the CLJFS research outcomes are quite trustworthy for identifying the emotions of preterm newborns from their cry signal.

5. CONCLUSION

In this research a novel CLJFS model is proposed for Cry-based Jaundiced Infant Signal (CLJFS) classification model. The crying is a newborn's primary communication, understanding its acoustic features can provide critical insights into the infant's condition. The proposed CLJFS model employs a multi-step process beginning with signal pre-processing using Stationary Wavelet Transform (SWT) for noise reduction and feature enhancement. Linear Prediction Coefficients (LPC) are extracted, followed by feature selection using the Least Absolute Shrinkage and Selection Operator (LASSO). A Spiking Neural Network (SNN) then categorizes the cries into three classes: hunger, fear, and discomfort. The effectiveness of the proposed CLJFS was evaluate using F1 score, accuracy, precision, recall, and specificity. The proposed CLJFS model achieved a classification accuracy 98.9%. The proposed model enhanced the total accuracy by 2.24%, 4.03%, 10.1%, and 7.03%, respectively. Future work will focus on expanding the dataset to include diverse cry signals from varied populations and integrating real-time monitoring systems. Additionally, exploring advanced neural architectures could further enhance the classification accuracy and robustness of the model.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

FUNDING STATEMENT

Authors did not receive any funding.

ACKNOWLEDGEMENTS

The author would like to express his heartfelt gratitude to the supervisor for his guidance and unwavering support during this research for his guidance and support.

REFERENCES

[1] Z. Khalilzad, A. Hasasneh, & C. Tadj, “Newborn cry-based diagnostic system to distinguish between sepsis and respiratory distress syndrome using combined acoustic features”, *Diagnostics*, vol. 12, no. 11, pp. 2802, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] P.D.R. Vincent, K. Srinivasan, C.Y. Chang, “Deep learning assisted premature infant cry classification via support vector machine models”, *Public Health Front*. vol. 9, pp. 670352, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] P. Rani, P. Kumar, V. Immanuel, P. Tharun, P. Rajesh, G. Fields, & G. D. Vaddeswaram, “Baby Cry Classification Using Machine Learning”, *Int. J. Innov. Sci. Res. Technol*, vol. 7, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] G. Coro, S. Bardelli, A. Cuttano, R. T. Scaramuzzo, & M. Ciantelli, “A self-training automatic infant-cry detector”, *Neural Computing and Applications*, vol. 35, no. 11, pp. 8543-8559, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] T. Jian, Y. Peng, W. Peng, & Z. Yang, “Research on LSTM+ Attention Model of Infant Cry Classification”, *J. Robotics Netw. Artif. Life*, vol. 8, no. 3, pp. 218-223, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] C. C. Onu, J. Lebensold, W. L. Hamilton, & D. Precup, “Neural transfer learning for cry-based diagnosis of perinatal asphyxia”, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] A. Gorin, C. Subakan, S. Abdoli, J. Wang, S. Latremouille, & C. Onu, “Self-supervised learning for infant cry analysis”, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] J. Li, M. Hasegawa-Johnson, & N. L. McElwain, “Analysis of acoustic and voice quality features for the classification of infant and mother vocalizations”, *Speech communication*, vol. 133, pp. 41-61, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] A. Gujral, K. Feng, G. Mandhyan, N. Snehil, & T. Chaspari, “Leveraging transfer learning techniques for classifying infant vocalizations”, *In 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI),* pp. 1-4, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] L. Zhang, C. P. Lim, Y. Yu, & M. Jiang, “Sound classification using evolving ensemble models and Particle Swarm Optimization”, *Applied soft computing*, vol. 116, pp. 108322, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] U. P. D. Pathirana, “An Intelligent hardware system for real-time infant cry detection and classification (Doctoral dissertation)”, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] F. S. Matikolaie, & C. Tadj, “Machine learning-based cry diagnostic system for identifying septic newborns”, *Journal of Voice*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] T. N. Maghfira, T. Basaruddin, & A. Krisnadhi, “Infant cry classification using cnn–rnn”, *In Journal of Physics: Conference Series IOP Publishing*, vol. 1528, no. 1, pp. 012019, 2020[CrossRef] [Google Scholar] [Publisher Link]

[14] M. Severini, D. Ferretti, E. Principi, & S. Squartini, “Automatic detection of cry sounds in neonatal intensive care units by using deep learning and acoustic scene simulation”, *IEEE Access*, vol. 7, 51982-51993, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[15] V. R. Joshi, K. Srinivasan, P. D. R. Vincent, V. Rajinikanth, & C. Y. Chang, “A Multistage Heterogeneous Stacking Ensemble Model for Augmented Infant Cry Classification”, *Frontiers in Public Health*, vol. 10, pp. 819865, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Y. C. Liang, I. Wijaya, M. T. Yang, J. R. Cuevas Juarez, & H. T. Chang, “Deep learning for infant cry recognition”, *International Journal of Environmental Research and Public Health*, vol. 19, no. 10, pp. 6311, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] S. Lahmiri, C. Tadj, C. Gargour, & S. Bekiros, Deep learning systems for automatic diagnosis of infant cry signals. Chaos, Solitons & Fractals, vol. 154, pp. 111700, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] J. Cha, & G. Bae, Deep learning based infant cry analysis utilizing computer vision. International Journal of Applied Engineering Research, vol. 17, no. 1, pp. 30-35, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] K. Ashwini, P. D. R. Vincent, K. Srinivasan, & C. Y. Chang, Deep learning assisted neonatal cry classification via support vector machine models. Frontiers in Public Health, 9, 2021. [CrossRef] [Google Scholar] [Publisher Link]

AUTHORS

J. Jency received a B.Sc. degree in Computer Science from Womens Christian College, Nagercoil, in 2015 and an M.Sc. degree in Computer Science from S. T. Hindu College, Nagercoil, in 2017. She received an M.Phil. degree in Computer Science from S.T. Hindu College, Nagercoil, in 2018. She is currently pursuing a Ph.D. degree in Computer Science at S.T. Hindu College, Nagercoil, Affiliated with Manonmaniam Sundaranar University, Tirunelveli, India. Her research interest includes digital image processing.

J. Anto Germin Sweeta She received the B.Sc. degree in Computer Science from Holy Cross College (Autonomous), Nagercoil, in 2016 and the M.Sc. degree in Computer Science from Scott Christian College, (Autonomous), Nagercoil, in 2018. She received the M.Phil. degree in Computer Science from S.T. Hindu College, Nagercoil, in 2019. She is currently pursuing the Ph.D. degree in Computer Science at Manonmaniam Sundaranar University, Tirunelveli. Her research interest includes digital image processing and business intelligence.

M. Annies Stelina received the B.Sc. degree in Information Technology from Nooral Islam Arts & Science, Kumaracoil, in 2007 and the Master degree in Computer Application from Sun College of Engineering and Technology, Erachakulam, in 2010. She received the M.Phil. degree in Computer Science from S.T. Hindu College, Nagercoil, in 2012. She is currently pursuing the Ph.D. degree in Computer Science at S.T. Hindu College, Nagercoil, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, India. Her research interest includes digital image processing.

Arrived: 12.11.2023
Accepted: 10.12.2023