Enhancing Autism Severity Classification: Integrating LSTM into CNNs for Multisite Meltdown Grading

Sumbul Alam[1]*, S Pravinth Raja[2], Yonis Gulzar[3]*, Mohammad Shuaib Mir[4]
Department of Computer Science Engineering, Presidency University, Bengaluru-560064, Karnataka, India[1, 2]
Department of Management Information Systems-College of Business Administration,
King Faisal University, Al-Ahsa 31982, Saudi Arabia[3, 4]

***Abstract*—Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by deficits in social interaction, verbal and non-verbal communication, and is often associated with cognitive and neurobehavioral challenges. Timely screening and diagnosis of ASD are crucial for early educational planning, treatment, family support, and timely medical intervention. Manual diagnostic methods are time-consuming and labor-intensive, underscoring the need for automated approaches to assist caretakers and parents. While various researchers have employed machine learning and deep learning techniques for ASD diagnosis, existing models often fall short in capturing the complexity of multisite meltdowns and fully leveraging the interdependence among these meltdowns for severity assessment in acquired facial images of children, hindering the development of a comprehensive grading system. This paper introduces a novel approach using a Long Short Term Memory (LSTM) integrated Convolution Neural Network (CNN) designed to identify multisite meltdowns and exploit their interdependence for severity assessment in ASD. The process begins with image pre-processing, involving discrete convolution filters for noise removal and contrast enhancement to improve image quality. The enhanced image then undergoes instance segmentation using the Segment Anything model to identify significant regions in the child's image. The segmented region is subjected to principal component analysis for feature extraction, and these features are utilized by the LSTM-integrated CNN for meltdown detection and severity classification. The model is trained using children's images extracted from videos, and testing is performed on videos captured during children's observations. Performance analysis reveals superior results, with a training accuracy of 88% and validation accuracy of 84%, outperforming conventional methods. This innovative approach not only enhances the efficiency of ASD diagnosis but also provides a more nuanced understanding of multisite meltdowns and their impact on severity, contributing to the development of a robust grading system.**

Keywords*—*Autism spectrum disorder; mutilating meltdown; convolution neural network; long short term memory; multisite meltdown; video classification; image classification

I. Introduction

Autism spectrum disorder (ASD) represents a complex neurodevelopmental syndrome characterized by a wide range of challenges in verbal and nonverbal communication skills, as well as behavioral and social interactions [1]. While ASD can manifest at any age, it typically becomes evident around the age of 2 or 3 when children start to withdraw, exhibit distinct behaviors, and present challenges in social engagement. The etiology of this disorder is diverse, and the underlying neurodevelopmental mechanisms are not fully understood [2].

The detection of a high degree of autism severity in a child is particularly concerning, as it often leads to the development of more frequent meltdowns. These meltdowns not only pose a risk of self-injury to the child but can also result in harm to caregivers or parents [3]. Early diagnosis of ASD is crucial, offering significant benefits in terms of intellectual development, adaptive behavior, and the reduction of overall severity. The advent of noninvasive acquisition technology has made disease diagnosis more feasible, but manual diagnosis remains a highly challenging and labor-intensive task. Consequently, there is a pressing need to develop an automated disease diagnosis tool for ASD.

In recent times, the pervasive influence of artificial intelligence (AI) has become increasingly apparent, bringing about transformative changes across a spectrum of fields and enriching various facets of our everyday existence [4,5]. It has redefined how we approach education [6], fine-tuned financial strategies [7], simplified agricultural workflows [8–16], and elevated healthcare diagnostics to new heights [17–23]. As it seamlessly integrates into these diverse sectors, AI continues to demonstrate its capacity for generating unparalleled efficiencies, refining decision-making procedures, and addressing intricate challenges with a precision derived from data-driven insights[24,25].

Researchers have turned to machine learning and deep learning approaches to enhance the accuracy of ASD diagnosis. These models leverage the concept of correlation and co-variation among spatio-temporal data. Specifically, Convolutional Neural Networks (CNNs) have proven highly capable of extracting features based on spatio-temporal descriptors to classify various gestures exhibited by ASD children. However, existing models have limitations, as they fail to compute the appearance of multisite meltdowns and fully exploit the interdependence among these meltdowns for severity computation in the acquired facial images of children, hindering the development of a comprehensive grading system [26].

This paper presents a novel approach, introducing a Long Short-Term Memory (LSTM) integrated Convolutional Neural Network designed to detect multisite meltdowns and exploit their interdependence for severity computation in ASD. The proposed methodology involves initial image pre-processing

*Correspondence: sumbulalam@presidencyuniversity.in (S.A); ygulzar@kfu.edu.sa (Y.G.)

using contrast enhancement and histogram equalization to improve image quality. The enhanced image is then subjected to an instance segmentation technique, termed the "segment anything" model, to identify significant regions in the child's image. These segmented regions undergo principal component analysis for feature extraction, which is then employed in the novel LSTM-integrated CNN [27] for meltdown detection and severity classification, achieving increased accuracy. The model is trained using images of children acquired from videos, and testing is conducted using videos acquired during children's observations.

The subsequent sections of the article are organized as follows: Section II provides a detailed problem statement and a review of the literature on detecting the meltdown state of autism spectrum disorder. Section III outlines the proposed deep learning methodology for detecting and classifying multisite meltdown states and their severity to establish a grading system. Section IV presents an experimental analysis of the proposed methodology on the disease dataset, including performance metrics such as accuracy through a confusion matrix. Finally, Section V concludes the work and offers future suggestions.

II. Related Work

In this part, numerous conventional approaches using machine learning and deep learning architecture to detect the autism spectrum disorder along meltdown state using behavioral and kinematics features has been detailed as follows:

A. Autism Spectrum Disorder Detection using Restricted Kinematic Features

In this literature, machine learning model is used to detect the Autism spectrum disorder using the restricted kinematic features from the kinematic data. In this computed on basis of movement during motor task. Entropy, amplitude, velocity, acceleration were considered as kinematic features and considered as symptom of the ASD. Machine learning classifier such as support vector machine were employed to classify or detect the ASD using the feature values which has cognitive flexibility and it has high manifestation [28].

B. Autism Spectrum Disorder Detection using Autoencoder-based Support Vector Machine - Recursive Feature Elimination Technique

In this literature, deep learning architecture is used to detect the ASD through functional connectivity features of the multiple regions. Functional connectivity feature is extracted and those features are employed to Recursive feature elimination technique to select the primitive features. Next, Autoencoder model is used to extract the high latent features and complicated features. It is considered as optimal features [29]. Those features were employed to softmax classifier which employs the Support vector machine to detect the ASD.

C. Autism Spectrum Disorder and Meltdown Detection on Facial Geometric Features using Recurrent Neural Network

In this literature, recurrent neural network is employed to detect and classify the autism spectrum disorder during the meltdown crisis. Initially model extracts the micro facial expression of children as geometric features and detects the child with autism or without autism. On detection of autism state , children is classified with severity of meltdown Hidden layer of the model process the feature to produce the optimal severity state of the children with meltdown [30].

D. Autism Spectrum Disorder and Meltdown Detection using Recurrent Attention Network on Morphological Features

In this literature, recurrent attention network is employed to detect and classify the autism spectrum disorder during the meltdown crisis. Initially model extracts the morphological features and detects the child with autism or without autism. On detection of autism state, children are classified with severity of meltdown. Attention layer of the model process the feature embedding to produce the optimal severity state of the children with meltdown with high efficiency [31].

E. Autism Spectrum Disorder and Meltdown Detection using Deep Neural Network on Audio-based Features

In this literature, deep neural network is employed to detect and classify the autism spectrum disorder during the meltdown crisis using audio based signals. Initially model extract the audio based features on transforming the speech data to mel spectrogram and those audio based feature like pitch, RMS and MFFS used to detect the child with autism or without autism. On detection of autism state, children are classified with severity of meltdown. Dense layer of the model process the features as embedding of features to produce the optimal severity state of the children with meltdown with high efficiency [32].

III. Proposed Model

In this section, we introduce a sophisticated system for grading multisite meltdowns and classifying the severity of autism spectrum disorder in children. Our approach integrates a Long Short-Term Memory (LSTM) with a Convolutional Neural Network (CNN) specifically tailored for assessing the intensity of meltdowns across various sites, encompassing expressions of distress such as crying, screaming, and stimming. The step-by-step processing to achieve this overarching objective is detailed below:

A. Image Preprocessing- Discrete Convolution Filter

This section outlines the application of noise reduction and contrast enhancement techniques to the acquired training images through the utilization of a discrete convolution filter. The initial phase of noise reduction in training images serves to eliminate blurriness, while the subsequent contrast enhancement enhances the overall image quality. The discrete convolution filter method, employed in this process, effectively heightens the sharpness of object edges within each image [33].

The discrete convolution filter method is instrumental in achieving this enhancement, producing a pronounced increase in edge sharpness. A special case of this method involves the averaging of brightness values, exemplified by the formula:

$$f(i,j)=w*h=\sum_{m=-a}^{a}\sum_{n=-b}^{b} w(m,n)h(i+m,j+n) \quad (1)$$

This Formula depicts a linear operation wherein the resulting value in the output image pixel f(i,j) is calculated as a linear combination of the brightness in a local neighborhood of the pixel h(i,j) in the input image. The function (w) in this context represents the convolution kernel, encapsulating the linear operations involved in this discrete convolution process.

B. Instance Segmentation - Segment Anything Model

The enhanced image undergoes application of the instance segmentation technique, known as the "Segment Anything" model, aimed at delineating significant regions within the child image. This model excels in segmenting every pixel with similar values, employing boundary probability algorithms (gPb and UCM) to calculate edge and its weight. A threshold is then applied for each pixel and its adjacent pixel to refine the segmentation process [34].

The instance segmentation process, as expressed by the following formula:

$$L(x,y)= \sum_{i=0}^{i<x,y<j}\sum_{j=0} I(i,j) \text{ where } 0<x<N \text{ and } 0<y<M \quad (2)$$

Where $0<x<N$ and $0<y<M$, involves aggregating pixel values within the specified range, providing a comprehensive representation of the preprocessed image. To integrate similar pixels within edges, a connected component approach is employed. The result is a hierarchical organization of regions obtained through a coarse-to-fine process.

C. Feature Extraction - Principle Component Analysis

The application of Principal Component Analysis (PCA) serves as a pivotal step in discerning the normal and meltdown states of children with autism. PCA, leveraging spatio-temporal information within each frame, identifies highly discriminating features. The execution of PCA results in a set of training data characterized by disconnectedness among data points and dense similarity within classes [35].

The analysis extends to examining segmented objects across two consecutive frames of images, calculating and defining features at various focal points such as the left eye, right eye, right mouth corner, left mouth corner, and nose. Each principal component represents the maximum variance among these focal points. To handle the complexity of computing features in high dimensions, PCA effectively minimizes dimensions without significant loss of feature information through matrix formation and distance calculation.

The composed feature vectors encapsulate facial interest components. Given an image of size $N \times N$, it is initially transformed into a 1D vector U, housing variance values of substantial magnitude. The variance for a specified feature X in an image is calculated by the formula:

$$\text{variance}(y) = \frac{\sum_{i=1}^{n} b(yi - y)\,(yi - y)}{n-1} \quad (3)$$

Furthermore, the covariance of features is calculated for the objects X and Y that change together with the mean, expressed as:

$$\text{Covariance}(y,x) = \frac{\sum_{i=1}^{n} a(yi - y)\,(xi - x)}{n-1} \quad (4)$$

The resulting Covariance Matrix, a $N \times N$ feature matrix, is represented by:

$$M_{ij} = \text{Covariance}(x,y) \quad (5)$$

D. Eigen Vector Analysis for Facial Feature Classification

The computation involves deriving the Eigen vector, denoted as M_{ij}^{Eigen}, serving as a feature vector that encapsulates principle feature groups. These feature values are accompanied by Eigen values and are pivotal for the subsequent classification of facial features.

For each meltdown and normal state, the associated feature values fall within the range of 0 to 1. Here, 0 signifies the normal state, while 1 designates the meltdown state. Table I presents the Eigen vector, composed of features extracted specifically for the meltdown state.

TABLE I. LIST OF FEATURE EXTRACTED FOR MELTDOWN STATE

Feature	Description
Eyes Closed	Distance between two eye lids
Mouth Open	Distance between two lips
Lips enlargement	Radius of the lips
Object in ears	Hand in the ears
Object in head	Hand in the head or hair

These features provide a detailed description of facial expressions during the meltdown state, including eye and mouth behaviors, lip enlargement, and hand placements indicative of distress. The Eigen vector, with its associated Eigen values, serves as a valuable tool for the effective classification of these facial features, contributing to a comprehensive understanding of autism states.

E. Long Short Term Memory Integrated Convolution Neural Network

The extracted features play a pivotal role in the functionality of the novel Long Short-Term Memory Integrated Convolutional Neural Network (LSTM-CNN) designed for the detection and classification of meltdown states, with a focus on assessing the severity. This innovative approach leverages hyperparameter-optimized layers to enhance the processing of information [36].

1) Long short term memory: The Long Short-Term Memory (LSTM) component within our integrated model is a key element endowed with the unique ability to learn and comprehend long-term dependencies through its intricate network connections. It excels in the storage of pertinent information within a memory cell while effectively discarding extraneous details. Each LSTM unit comprises a memory cell equipped with two gates: input and output, and a forget gate. In the context of our research, the LSTM model serves as a repository for multisite meltdown feature maps, a product of the convolution and max-pooling layers in the Convolutional Neural Network (CNN). These feature maps encapsulate crucial patterns essential for the accurate detection and classification of meltdown states.

In order to fine-tune the performance of our LSTM-CNN model, we employ hyperparameter optimization, as outlined in Table II.

TABLE II. HYPERPARAMETER TUNING

Hyperparameter	Value
Learning rate	10^{-6}
Epoch Value	100
Activation function	ReLu
Loss Function	Cross Entropy

The hyperparameters, meticulously chosen and specified in Table II, play a pivotal role in shaping the learning dynamics of our model. The learning rate, set at 10^-6, determines the step size during optimization, ensuring a balance between accuracy and efficiency. The epoch value of 100 signifies the number of times the entire dataset is processed during training, influencing the model's convergence. The ReLU activation function is employed to introduce non-linearity, enhancing the model's capacity to learn intricate patterns. Finally, the Cross Entropy loss function measures the dissimilarity between predicted and actual values, guiding the model towards optimal performance. This comprehensive hyperparameter tuning aims to maximize the LSTM-CNN model's efficacy in detecting and classifying meltdown states with a nuanced understanding of their severity.

2) Convolution neural network: The Convolutional Neural Network (CNN) serves as a cornerstone in our methodology, tasked with processing the intricately extracted features across multiple layers. Its primary objective is to adeptly detect and classify the multisite meltdown state, discerning the severity level inherent in each case. The CNN's architecture is meticulously designed, incorporating essential elements such as the convolution layer, max-pooling layer, and fully connected layer.

The convolution layer plays a critical role in feature extraction, employing filters to scan and identify distinctive patterns within the input data. This process enables the CNN to capture essential spatial hierarchies and dependencies in the multisite meltdown features. Subsequently, the max-pooling layer strategically downsizes the spatial dimensions of the extracted features, promoting computational efficiency and reducing the risk of overfitting.

The fully connected layer, a crucial component of the CNN architecture, is responsible for processing linear features extracted from the preceding layers. It incorporates an activation function to introduce non-linearity, allowing the model to learn complex relationships within the data. Furthermore, the softmax function within the fully connected layer serves a dual purpose – detection and classification. It assigns probabilities to different meltdown states, facilitating a nuanced understanding of the severity levels associated with each classification.

To guide the training process effectively, a loss function is integrated into the fully connected layer. This function calculates the classification error, providing feedback to the model during the training phase. The objective is to minimize this error, enhancing the CNN's ability to accurately detect and classify multisite meltdown states.

For a visual representation of our proposed model's architecture, refer to Fig. 1. This diagram encapsulates the intricate interplay between the convolution layer, max-pooling layer, and fully connected layer, offering a comprehensive overview of the network's structure and functionality. The synergy of these components within the CNN underscores its efficacy in robustly addressing the detection and classification challenges posed by multisite meltdown scenarios.

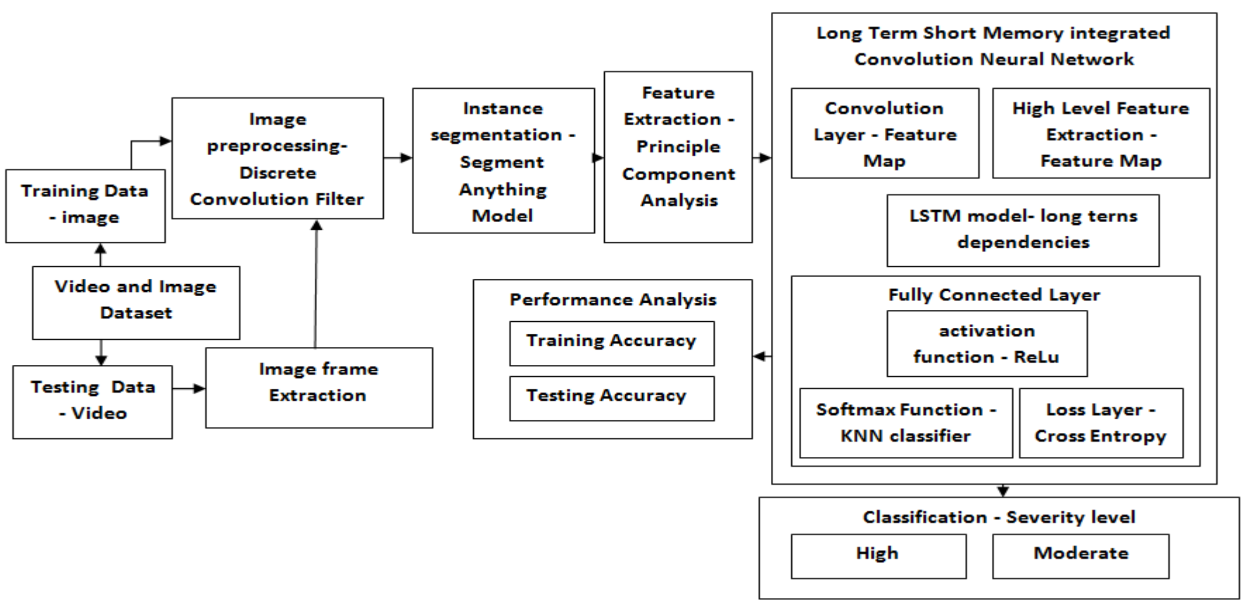

Fig. 1. Proposed architecture.

- Convolution Layer: The Convolution Layer is a pivotal element in our architecture, containing multiple filters and kernels that convolve with features extracted from different regions. This convolution operation produces a feature map that represents the underlying patterns in the meltdown state. Mathematically, convolution involves the multiplication of the feature vector, containing information about the meltdown state in a specified region, with multiple filters [37]. The convolution process is expressed as in the Formula:

$$x_n = \sum_{K=0}^{N-1} Y_K F_{n-k} \tag{6}$$

Here, *Y* represents the feature, and *F* is the filter.

The convolution layer generates a feature map through convolution operations, encompassing both low-level and latent features. The convergence of this feature map is facilitated by epochs, incrementally increasing feature generation. Normalization through the Rectified Linear Unit (ReLU) activation function further refines the feature map, obtaining a linear representation. The cosine distance measure is then employed to compute the distance among features.

- Pooling Layer: The Pooling Layer follows the convolutional operations, serving to further reduce the features obtained from the convolution layer. This step is crucial for high-level meltdown feature extraction and is essentially a form of down-sampling, diminishing the dimensions of facial features and retaining only selected weighted meltdown features. The Max Pooling layer plays a vital role in connecting the meltdown features into small patches, estimating the maximum number of features for each subset. This process enhances model generalization [38].

- Long Short Term Memory Layer: The LSTM Layer is employed to store the feature map derived from the convolution and max-pooling layers of the Convolutional Neural Network. It excels in preserving long-term dependencies through its intricate network connections. The stored meltdown features, each assigned a weight value, reside in the memory cell of each LSTM unit. These features, converted into a feature matrix, are input into the LSTM for the fusion of multisite meltdown information [39].

$$C_t = \tanh(X_t * V_t + H_{t-1} * W_t) \tag{7}$$

In the CNN-LSTM hybrid model, normal and meltdown features are extracted from both the convolutional and LSTM layers. The ordering of these features is then utilized, where *Ht* represents cell memory information, and *Wt* represents the weight vector.

- Dense Layer – Fully Connected Layer: The Dense Layer, organized as a fully connected layer, processes the feature map composed of multisite meltdown features across facial regions. This layer extracts discriminative features related to crying, stimming, and screaming. The activation function is applied for feature normalization and flattening, addressing non-linearity and overfitting concerns in the feature maps.

- Softmax Layer: The Softmax Layer, integrated into the fully connected layer, is crucial for detecting the meltdown state. It combines these states, assigning aggregate weights to identify the severity of multisite meltdown states using a Naive Bayes classifier. A loss layer is further incorporated to minimize feature variance across classes. The Softmax function, as represented in Formula (8), calculates the probability distribution:

$$\text{Softmax Function } P_j = \frac{e^{x_j}}{\sum_1^k e^{x_k}} \tag{8}$$

where, e^{xj} is the feature map long dependency vector.

- Classifier and Decision Rule: The feature vector is projected for classification by applying Bayes theorem, utilizing a maximum likelihood function to aggregate similar emotion features of autistic children based on feature values. The final classification result is generated by integrating results according to the decision rule. Feature values related to crying, screaming, and stimming form distinct classes. The Maximum Likelihood function, as given in Eq.9, incorporates the class, feature values, and density function:

$$f_n(y, \theta) = \prod_{k=1}^{n} f_k \tag{9}$$

Here, Y is the class of the meltdown, θ is the vector containing feature values, and f_k is the density function. This comprehensive approach ensures a robust and nuanced classification of multisite meltdown states based on their severity levels.

Algorithm 1: Multisite Meltdown Detection and Severity Classification.
Input : Video and Images of the Child observations Output: Detection of Multisite Meltdown and severity classes Process Train() Preprocessing of training images () Contrast Enhancement () Preprocessed image =Discrete Convolution filter (Training images) Instance Segmentation() Segment = Segment Anything Model(Preprocessed image) Feature extraction() Transform the image pixel of segment into matrix Compute the covariance and correlation on the matrix Determine the eigen value and eigen vector Eigen vector = Feature Vector Disease detection and classification () Convolution Neural Network() Convolution Layer () = VGG19() Low level feature = Kernel (Feature vector) Feature map = ReLu(Low level Features) Max pooling layer () High level feature = Kernel (Feature vector) Feature map = ReLu(Low level Features) LTSM layer ()

Store feature map as long terms dependencies
Feature Dependencies= combining the multiple meltdown
Fully connected layer ()
Activation function = ReLu()
Softmax function = Naive Bayes (Feature dependencies map)
Detection of disease = { Crying , Screaming , Stimming}
Severity Class= { High , Moderate }
Loss function - Cross Entropy()

The algorithm offers a holistic approach to Multisite Meltdown Detection and Severity Classification. It integrates preprocessing, segmentation, feature extraction, and a robust combination of Convolution Neural Network and Long Short-Term Memory for accurate and nuanced results. The inclusion of a variety of techniques, such as contrast enhancement, instance segmentation, and feature mapping, contributes to the algorithm's effectiveness in handling complex scenarios related to autism severity classification. The architecture demonstrates a deep understanding of both low-level and high-level features, providing a comprehensive solution for the challenging task at hand.

IV. Experimental Results

In this pivotal section, we delve into a thorough analysis of the experimental outcomes, leveraging cross-fold validation on a simulated dataset within the Python environment [40]. The performance evaluation of our proposed architecture for autism severity classification is meticulously conducted, with optimal parameters defining the model's performance. The implementation utilizes the versatile Scikit-learn package, incorporating various machine learning algorithms, and OpenCV for efficient image processing and preparation.

A. Dataset Description - Meltdown Crisis

The cornerstone of our investigation lies in the Meltdown Crisis dataset, a robust collection designed for identifying multisite meltdown severity in autism children. Comprising 59 videos, this dataset offers a detailed narrative, encompassing facial expressions and physical activities during the meltdown crisis. The dataset is structured into emotion frames and non-emotion frames, categorizing various states such as normal, post-crisis, and meltdown crisis states. For streamlined evaluation, the dataset is strategically partitioned into training, testing, and validation sets [41].

The training phase involves the utilization of images extracted from children in the videos, while testing is conducted on videos observed during children's activities. Fig. 2 presents the confusion matrix for the validation dataset, consisting of 59 videos. This visual representation encapsulates the model's proficiency in classifying instances across different categories.

Fig. 3 provides a comprehensive snapshot of the training and validation accuracy of our model. This graphical representation elucidates the learning trajectory of the model over multiple epochs. The model achieves commendable results, boasting an 88% training accuracy and an 84% validation accuracy, reflecting its robust learning capabilities.

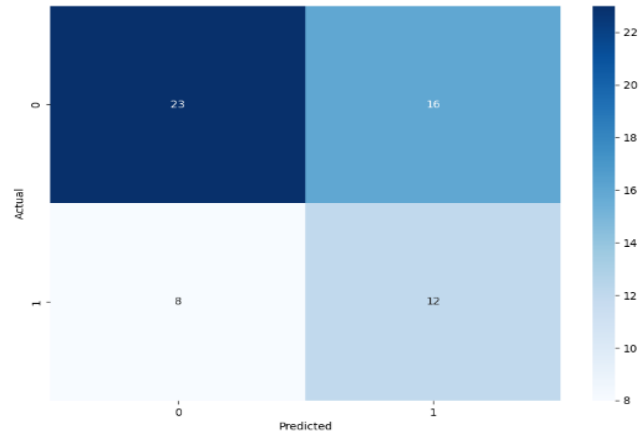

Fig. 2. Confusion matrix.

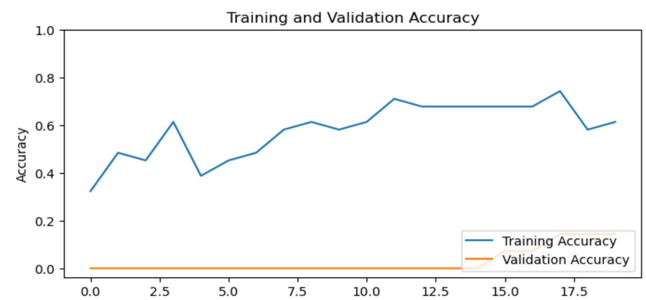

Fig. 3. Training and validation accuracy of the model.

Performance analysis extends to the examination of the training and validation loss, as illustrated in Fig. 4. These curves offer insights into the model's optimization process, demonstrating a balanced and decreasing trend in error reduction over the course of training and validation.

Fig. 4. Training and validation of the model.

TABLE III. Performance Evaluation of the Model

Technique	Accuracy		Loss	
	Training	**Validation**	**Training**	**Validation**
CNN+LSTM	88	84	0.8	0.1

Table III meticulously encapsulates the quantitative results of our model. The training accuracy, validation accuracy, training loss, and validation loss are presented, offering a detailed performance snapshot. The CNN–LSTM models were meticulously trained for 100 epochs, employing a batch size of 128, and utilizing a cross-entropy loss function. These parameters were strategically chosen to ensure a robust and effective training process.

The proposed model exhibits superior performance when benchmarked against conventional approaches. With a training accuracy of 88% and a validation accuracy of 84%, coupled with minimal training and validation loss, our CNN–LSTM model demonstrates efficacy in autism severity classification. These findings underscore the potential of our integrated architecture in providing nuanced and accurate assessments in the context of multisite meltdown grading.

V. Conclusion

In this study, we introduced a novel approach for detecting multisite meltdowns in children with Autism Spectrum Disorder (ASD) using a Long Short-Term Memory (LSTM) [42] integrated Convolutional Neural Network (CNN). Our designed architecture aims to leverage the dependency among multisite meltdowns to enhance the severity computation, ultimately contributing to the development of a robust grading system. The comprehensive pipeline of our model encompasses pre-processing techniques to enhance image quality, the utilization of the Segment Anything model for segmentation, and principle component analysis for feature extraction. These steps are crucial in isolating significant regions within child images and extracting pertinent features.

The extracted features are subsequently fed into the CNN+LSTM classifier, which effectively detects and classifies multisite meltdowns, providing valuable insights into their severity. The model's performance analysis yielded promising results, with a training accuracy of 88% and a validation accuracy of 84%. This underscores the efficacy of our proposed architecture in accurately identifying and grading multisite meltdowns in children with ASD.

While our current model has shown promising results, there are avenues for future research and improvement. Firstly, the inclusion of a larger and more diverse dataset could enhance the model's generalization capabilities. Exploring advanced techniques for feature extraction and segmentation may further refine the model's ability to capture subtle nuances in facial expressions during meltdowns.

Additionally, investigating real-time applications and deployment in clinical settings could provide valuable insights into the model's practical utility. Fine-tuning hyperparameters and exploring alternative neural network architectures may also contribute to optimizing the model's performance.

Acknowledgment

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under the Project GRANT5, 306.

References

[1] Alam, S.; Raja, P.; Gulzar, Y. Investigation of Machine Learning Methods for Early Prediction of Neurodevelopmental Disorders in Children. Wirel Commun Mob Comput 2022, 2022.

[2] Guha, T.; Yang, Z.; Ramakrishna, A.; Grossman, R.B.; Darren, H.; Lee, S.; Narayanan, S.S. On Quantifying Facial Expression-Related Atypicality of Children with Autism Spectrum Disorder. Proc IEEE Int Conf Acoust Speech Signal Process 2015, 2015, 803–807, doi:10.1109/ICASSP.2015.7178080.

[3] Guo, J.; Zhou, S.; Wu, J.; Wan, J.; Zhu, X.; Lei, Z.; Li, S.Z. Multi-Modality Network with Visual and Geometrical Information for Micro Emotion Recognition. ieeexplore.ieee.orgJ Guo, S Zhou, J Wu, J Wan, X Zhu, Z Lei, SZ Li2017 12th IEEE international conference on automatic face, 2017•ieeexplore.ieee.org.

[4] Ayoub, S.; Gulzar, Y.; Reegu, F.A.; Turaev, S. Generating Image Captions Using Bahdanau Attention Mechanism and Transfer Learning. Symmetry (Basel) 2022, 14, 2681.

[5] Hamid, Y.; Elyassami, S.; Gulzar, Y.; Balasaraswathi, V.R.; Habuza, T.; Wani, S. An Improvised CNN Model for Fake Image Detection. International Journal of Information Technology 2023, 15, 5–15, doi:10.1007/S41870-022-01130-5.

[6] Sahlan, F.; Hamidi, F.; Misrat, M.Z.; Adli, M.H.; Wani, S.; Gulzar, Y. Prediction of Mental Health Among University Students. International Journal on Perceptive and Cognitive Computing 2021, 7, 85–91.

[7] Gulzar, Y.; Alwan, A.A.; Abdullah, R.M.; Abualkishik, A.Z.; Oumrani, M. OCA: Ordered Clustering-Based Algorithm for E-Commerce Recommendation System. Sustainability 2023, Vol. 15, Page 2947 2023, 15, 2947, doi:10.3390/SU15042947.

[8] Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. Sustainability 2023, 15, 1906.

[9] Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. Sustainability 2023, 15, 901.

[10] Dhiman, P.; Kaur, A.; Balasaraswathi, V.R.; Gulzar, Y.; Alwan, A.A.; Hamid, Y. Image Acquisition, Preprocessing and Classification of Citrus Fruit Diseases: A Systematic Literature Review. Sustainability 2023, Vol. 15, Page 9643 2023, 15, 9643, doi:10.3390/SU15129643.

[11] Albarrak, K.; Gulzar, Y.; Hamid, Y.; Mehmood, A.; Soomro, A.B. A Deep Learning-Based Model for Date Fruit Classification. Sustainability 2022, 14.

[12] Hamid, Y.; Wani, S.; Soomro, A.B.; Alwan, A.A.; Gulzar, Y. Smart Seed Classification System Based on MobileNetV2 Architecture. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT); IEEE, 2022; pp. 217–222.

[13] Gulzar, Y.; Hamid, Y.; Soomro, A.B.; Alwan, A.A.; Journaux, L. A Convolution Neural Network-Based Seed Classification System. Symmetry (Basel) 2020, 12, 2018.

[14] Gulzar, Y.; Ünal, Z.; Akta¸s, H.A.; Mir, M.S. Harnessing the Power of Transfer Learning in Sunflower Disease Detection: A Comparative Study. Agriculture 2023, Vol. 13, Page 1479 2023, 13, 1479, doi:10.3390/AGRICULTURE13081479.

[15] Malik, I.; Ahmed, M.; Gulzar, Y.; Baba, S.H.; Mir, M.S.; Soomro, A.B.; Sultan, A.; Elwasila, O. Estimation of the Extent of the Vulnerability of Agriculture to Climate Change Using Analytical and Deep-Learning Methods: A Case Study in Jammu, Kashmir, and Ladakh. Sustainability 2023, Vol. 15, Page 11465 2023, 15, 11465, doi:10.3390/SU151411465.

[16] Aggarwal, S.; Gupta, S.; Gupta, D.; Gulzar, Y.; Juneja, S.; Alwan, A.A.; Nauman, A. An Artificial Intelligence-Based Stacked Ensemble Approach for Prediction of Protein Subcellular Localization in Confocal Microscopy Images. Sustainability 2023, Vol. 15, Page 1695 2023, 15, 1695, doi:10.3390/SU15021695.

[17] Gulzar, Y.; Khan, S.A. Skin Lesion Segmentation Based on Vision Transformers and Convolutional Neural Networks—A Comparative Study. Applied Sciences 2022, Vol. 12, Page 5990 2022, 12, 5990, doi:10.3390/APP12125990.

[18] Mehmood, A.; Gulzar, Y.; Ilyas, Q.M.; Jabbari, A.; Ahmad, M.; Iqbal, S. SBXception: A Shallower and Broader Xception Architecture for Efficient Classification of Skin Lesions. Cancers 2023, Vol. 15, Page 3604 2023, 15, 3604, doi:10.3390/CANCERS15143604.

[19] Khan, F.; Ayoub, S.; Gulzar, Y.; Majid, M.; Reegu, F.A.; Mir, M.S.; Soomro, A.B.; Elwasila, O. MRI-Based Effective Ensemble Frameworks for Predicting Human Brain Tumor. Journal of Imaging 2023, Vol. 9, Page 163 2023, 9, 163, doi:10.3390/JIMAGING9080163.

[20] Majid, M.; Gulzar, Y.; Ayoub, S.; Khan, F.; Reegu, F.A.; Mir, M.S.; Jaziri, W.; Soomro, A.B. Enhanced Transfer Learning Strategies for Effective Kidney Tumor Classification with CT Imaging. International

Journal of Advanced Computer Science and Applications 2023, 14, 2023, doi:10.14569/IJACSA.2023.0140847.

[21] Anand, V.; Gupta, S.; Gupta, D.; Gulzar, Y.; Xin, Q.; Juneja, S.; Shah, A.; Shaikh, A. Weighted Average Ensemble Deep Learning Model for Stratification of Brain Tumor in MRI Images. Diagnostics 2023, Vol. 13, Page 1320 2023, 13, 1320, doi:10.3390/DIAGNOSTICS13071320.

[22] Majid, M.; Gulzar, Y.; Ayoub, S.; Khan, F.; Reegu, F.A.; Mir, M.S.; Jaziri, W.; Soomro, A.B. Using Ensemble Learning and Advanced Data Mining Techniques to Improve the Diagnosis of Chronic Kidney Disease. International Journal of Advanced Computer Science and Applications 2023, 14, doi:10.14569/IJACSA.2023.0141050.

[23] Khan, S.A.; Gulzar, Y.; Turaev, S.; Peng, Y.S. A Modified HSIFT Descriptor for Medical Image Classification of Anatomy Objects. Symmetry (Basel) 2021, 13, 1987.

[24] Dhiman, P.; Bonkra, A.; Kaur, A.; Gulzar, Y.; Hamid, Y.; Mir, M.S.; Soomro, A.B.; Elwasila, O. Healthcare Trust Evolution with Explainable Artificial Intelligence: Bibliometric Analysis. Information 2023, Vol. 14, Page 541 2023, 14, 541, doi:10.3390/INFO14100541.

[25] Hanafi, M.F.F.M.; Nasir, M.S.F.M.; Wani, S.; Abdulghafor, R.A.A.; Gulzar, Y.; Hamid, Y. A Real Time Deep Learning Based Driver Monitoring System. International Journal on Perceptive and Cognitive Computing 2021, 7, 79–84.

[26] Masmoudi, M.; … S.J.-2019 15th I.; 2019, undefined Meltdowncrisis: Dataset of Autistic Children during Meltdown Crisis. ieeexplore.ieee.orgM Masmoudi, SK Jarraya, M Hammami2019 15th International Conference on Signal-Image Technology, 2019•ieeexplore.ieee.org.

[27] Haweel, R.; Dekhil, O.; Shalaby, A.; Mahmoud, A.; Ghazal, M.; Khalil, A.; Keynton, R.; Barnes, G.; El-Baz, A. A Novel Framework for Grading Autism Severity Using Task-Based FMRI. Proceedings - International Symposium on Biomedical Imaging 2020, 2020-April, 1404–1407, doi:10.1109/ISBI45749.2020.9098430.

[28] Herringshaw, A.J.; Ammons, C.J.; DeRamus, T.P.; Kana, R.K. Hemispheric Differences in Language Processing in Autism Spectrum Disorders: A Meta-Analysis of Neuroimaging Studies. Autism Research 2016, 9, 1046–1057, doi:10.1002/AUR.1599.

[29] Redcay, E.; Courchesne, E. Deviant Functional Magnetic Resonance Imaging Patterns of Brain Activity to Speech in 2-3-Year-Old Children with Autism Spectrum Disorder. Biol Psychiatry 2008, 64, 589–598, doi:10.1016/j.biopsych.2008.05.020.

[30] Jarraya, S.K.; Masmoudi, M.; Hammami, M. Compound Emotion Recognition of Autistic Children During Meltdown Crisis Based on Deep Spatio-Temporal Analysis of Facial Geometric Features. IEEE Access 2020, 8, 69311–69326, doi:10.1109/access.2020.2986654.

[31] Ke, F.; Yang, R. Classification and Biomarker Exploration of Autism Spectrum Disorders Based on Recurrent Attention Model. IEEE Access 2020, 8, 216298–216307, doi:10.1109/access.2020.3038479.

[32] Eni, M.; Dinstein, I.; Ilan, M.; Menashe, I.; Meiri, G.; Zigel, Y. Estimating Autism Severity in Young Children From Speech Signals Using a Deep Neural Network. IEEE Access 2020, 8, 139489–139500, doi:10.1109/access.2020.3012532.

[33] Liu, W.; Li, M.; Yi, L. Identifying Children with Autism Spectrum Disorder Based on Their Face Processing Abnormality: A Machine Learning Framework. Autism Research 2016, 9, 888–898, doi:10.1002/aur.1615.

[34] Rudovic, O.; Utsumi, Y.; Lee, J.; Hernandez, J.; Ferrer, E.C.; Schuller, B.; Picard, R.W. CultureNet: A Deep Learning Approach for Engagement Intensity Estimation from Face Images of Children with Autism. IEEE International Conference on Intelligent Robots and Systems 2018, 339–346, doi:10.1109/IROS.2018.8594177.

[35] Lombardo, M. V; Pramparo, T.; Gazestani, V.; Warrier, V.; Bethlehem, R.A.I.; Carter Barnes, C.; Lopez, L.; Lewis, N.E.; Eyler, L.; Pierce, K.; et al. Large-Scale Associations between the Leukocyte Transcriptome and BOLD Responses to Speech Differ in Autism Early Language Outcome Subtypes. Nat Neurosci 2018, 21, 1680–1688, doi:10.1038/s41593-018-0281-3.

[36] Lord, C.; Elsabbagh, M.; Baird, G.; Veenstra-Vanderweele, J. Autism Spectrum Disorder. Lancet 2018, 392, 508–520, doi:10.1016/S0140-6736(18)31129-2.

[37] Gotham, K.; Pickles, A.; Lord, C. Standardizing ADOS Scores for a Measure of Severity in Autism Spectrum Disorders. J Autism Dev Disord 2009, 39, 693–705, doi:10.1007/s10803-008-0674-3.

[38] Stoner, R.; Chow, M.L.; Boyle, M.P.; Sunkin, S.M.; Mouton, P.R.; Roy, S.; Wynshaw-Boris, A.; Colamarino, S.A.; Lein, E.S.; Courchesne, E. Patches of Disorganization in the Neocortex of Children with Autism. N Engl J Med 2014, 370, 1209–1219, doi:10.1056/NEJMoa1307491.

[39] Karten, A.; Hirsch, J. Brief Report: Anomalous Neural Deactivations and Functional Connectivity during Receptive Language in Autism Spectrum Disorder: A Functional MRI Study. J Autism Dev Disord 2015, 45, 1905–1914, doi:10.1007/s10803-014-2344-y.

[40] Kliuev, E.A.; Sheyko, G.E.; Dunayev, M.G.; Abramov, S.A.; Dvoryaninova, V. V; Balandina, O. V; Karyakin, N.N.; Belova, A.N. The Role of Functional MRI in Understanding the Origin of Speech Delay in Autism Spectrum Disorders. Sovremennye tehnologii v medicine 2019, 11, 66, doi:10.17691/stm2019.11.3.09.

[41] Ayoub, S.; Gulzar, Y.; Rustamov, J.; Jabbari, A.; Reegu, F.A.; Turaev, S. Adversarial Approaches to Tackle Imbalanced Data in Machine Learning. Sustainability 2023, Vol. 15, Page 7097 2023, 15, 7097, doi:10.3390/SU15097097.

[42] Khan, F.; Gulzar, Y.; Ayoub, S.; Majid, M.; Mir, M.S.; Soomro, A.B. Least Square-Support Vector Machine Based Brain Tumor Classification System with Multi Model Texture Features. Front Appl Math Stat 2023, 9, 1324054, doi:10.3389/FAMS.2023.1324054.