IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 December 2023, accepted 25 January 2024, date of publication 29 January 2024, date of current version 5 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3359533

RESEARCH ARTICLE

Identification of High Energy Gamma Particles From the Cherenkov Gamma Telescope Data Using a Deep Learning Approach

K. KARTHICK[1], (Member, IEEE), S. AKILA AGNES[2], S. SENDIL KUMAR[3], SULTAN ALFARHOOD[4], AND MEJDL SAFRAN[4]

[1]Department of Electrical and Electronics Engineering, GMR Institute of Technology, Rajam, Andhra Pradesh 532127, India
[2]Department of Computer Science and Engineering, GMR Institute of Technology, Rajam, Andhra Pradesh 532127, India
[3]Department of Electrical and Electronics Engineering, S. A. Engineering College (Autonomous), Chennai 600077, India
[4]Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding authors: K. Karthick (karthick.k@gmrit.edu.in) and Sultan Alfarhood (sultanf@ksu.edu.sa)

This research is funded by the Researchers Supporting Project Number (RSPD2024R890), King Saud University, Riyadh, Saudi Arabia.

ABSTRACT Atmospheric Cherenkov telescopes have enabled recent breakthroughs in gamma-ray astronomy, enabling the study of high-energy gamma particles in over 90 galactic and extragalactic regions. The significance of this work arises from the complexity of the data captured by the telescope. Traditional methods may struggle to effectively distinguish between gamma (signal) and hadron (background) events, due to intricate temporal relationships inherent in the data. The dataset used for this research, sourced from the UCI ML repository, simulates the registration of gamma particles. The challenge is to develop a classification model that accurately identifies these gamma events while handling inherent data complexities and normalizing skewed distributions. To address this challenge, a classification model is developed using ten features from the MAGIC gamma telescope dataset. This research introduces the innovative application of deep learning techniques, specifically Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (Bi-LSTM), to the field of gamma-ray astronomy to classify high-energy gamma particles detected by the Atmospheric Cherenkov telescopes. Furthermore, the research introduces the application of square root transformation as a method to address skewness and kurtosis in the dataset. This preprocessing technique aids in normalizing data distributions, which is crucial for accurate model training and classification. By leveraging the power of deep learning and innovative data transformations, the best accuracy of 88.71% is achieved by the LSTM+ReLU model with three layers for gamma and hadron particle classification. These findings offer insights into fundamental astrophysical processes and contribute to the advancement of gamma-ray astronomy.

INDEX TERMS Deep learning, gamma-rays, gamma-ray telescopes, LSTM, signal classification.

I. INTRODUCTION

Ground-based very high energy (VHE) gamma ray astronomy is one of the youngest entries among the numerous branches of astronomy. This field was pioneered by the Whipple group, who detected the first TeV gamma rays from the Crab Nebula in 1989 [1]. Gamma-ray observation utilising the imaging atmospheric Cherenkov technique (IACT) from 100 GeV to 10 TeV has grown rapidly in recent years. Besides the charged particles of cosmic radiation, cosmic gamma-ray photons are not affected by magnetic fields. As a result, studying gamma rays enables us to understand the characteristics of the sources and the acceleration mechanisms of cosmic rays [2]. Since the gamma rays are uncharged, they can travel in a straight line and hence carry a signature of the astrophysical source's original path. This

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia.

feature of gamma rays makes them among the most valuable tools for studying not only the source location but also for understanding and thereby unravelling the physical mechanisms at work inside the astrophysical laboratory. Gamma ray investigations can also provide information regarding dark matter annihilation [3]. Leptonic and hadronic particles can both generate gamma rays.

Because gamma rays are absorbed in the atmosphere, studying the universe using gamma rays from Earth is extremely difficult [4]. Prior to the successful adoption of ground-based detection of gamma rays, satellites were the only means of exploring the high energy universe [5]. The most significant constraint on the lower sensitivity of satellite-based investigations is their confined effective area. As long as there was no significant improvement in the effective area, there was very little hope of efficiently probing the VHE universe. This goal was achieved through the successful operation of ground-based telescopes detecting and identifying atmospheric Cherenkov radiation [6]. Ground-based telescopes have an effective area of $\sim 10^5$ m^2 as compared to ~ 1 m^2 for satellite-based telescopes. Despite the fact that the gamma ray to hadron event ratio is $\sim 10^{-3}$, gamma-hadron classification is a major issue. Furthermore, adopting a stereoscopic system with two or more telescopes enhances the rejection of hadronic events by a factor of 100 [7]. As high-energy cosmic rays pass through the atmosphere, they produce secondary particles, resulting in widespread air showers. These particles travel at relativistic speeds, resulting in Cherenkov radiation in the atmosphere. This radiation is detected using ground-based detectors. The IACT-based telescopes gather the Cherenkov photons that fall on these telescopes. These telescopes have a camera that collects Cherenkov photons that are reflected from the mirror. The camera is made up of photomultiplier tubes that are linked by rapid electronics that digitize and observe the Cherenkov photon pulse [8].

Deep learning methods [9], [10], [11], [12] provide better solutions for multivariate problems. Deep Learning is effective due to its superior accuracy when handling large amounts of data. They are very good at processing visual, speech, and text data [13], [14], [15], [16].

RNN (Recurrent Neural Network) is well suited for sequential data analysis tasks, such as time series or language data. In the context of the MAGIC gamma telescope dataset classification, RNN has been used to take into account the temporal relationship between the features. The sequence of features extracted from the telescope camera for each event can be considered as a sequence of input vectors. By using RNN, the information from the previous time steps can be retained and used to help classify the current input. This is particularly useful in scenarios where the sequence of features is important for making accurate predictions, and where traditional machine learning algorithms may not be able to capture this information. Therefore, by using RNNs, the model has been taken into account the sequence of features and improve the accuracy of classification.

The main objective of this research study is to develop a deep learning-based classification model by adopting the MAGIC gamma telescope dataset from the UCI machine learning (ML) repository. Multidimensional datasets are extremely challenging to manage using traditional methods. The standard for signal characterization in ground-based atmospheric Cherenkov devices comprises multidimensional data. In this article, a classification model based on deep learning have been employed to sort gamma and hadron signals from MAGIC gamma telescope data.

II. RELATED WORKS

Zhang et al. [17] made an effective method for extracting parameters from gamma-ray emissions of special nuclear materials (SNM) and identifying SNM classes using a back-propagation neural network (BPNN) and template matching approach.

For a ground-based atmospheric Cherenkov telescope, Mradul et al. [18] looked into gamma-hadron separation in great detail. They employed Monte Carlo event simulation to evaluate and compare various supervised ML techniques. These included the Random Forest (RF) method, Artificial Neural Networks (ANN), Linear Discriminant analysis, Naive Bayes (NB) Classifiers, Support Vector Machines (SVM), and the conventional dynamic supercut technique. For gamma-hadron segregation, the Random Forest approach has proved to be the most effective ML method.

A case study that compares multivariate classification approaches was presented by Bock et al. [19]. The input for the imaging gamma-ray Cherenkov telescope consists of Monte Carlo data, which has been generated and preprocessed. Both incoming gamma rays and hadronic showers contribute to this information, which is then separated into its respective categories. The data is well-suited for testing classification algorithms due to the low contrast between the signal (gamma) and background (hadrons).

Hirashima et al. [20] utilized a ML approach along with novel features, including 3D dosiomics characteristics coupled with plan and dosiomics features, to estimate and categorize the gamma passing rate value for volumetric modulated arc therapy (VMAT) plans. Using the high-energy stereoscopic system (HESS), Ohm et al. [21] demonstrated the stability and ability to minimize background noise of their tree classification algorithm compared to the HESS standard analysis.

Very Energetic Radiation Imaging Telescope Array System (VERITAS) data was classified using "boosted decision trees" by Krause et al. [22], and the results indicated improved sensitivity compared to the usual VERITAS analysis.

To do this, Brill et al. [23] presented a combination of CNNs and RNNs. Using CTLearn, a freely available Python programme that use deep learning to analyse data from IACTs, the team created a CNN-RNN network and discovered inadequate evidence that ordering telescope pictures by cumulative magnitude improves background rejection performance. Nieto et al. [24] employed CTLearn, a python package that uses deep learning in order to analyse data from IACT arrays.

Using the water Cherenkov detector with a smaller water volume and four PMTs and analysing the PMT signal's spatial and time patterns with a ML method enables optimum muon tagging, as stated by Conceiço et al. [25].

Algorithms are the backbone of ML, which examines data, learns from it, and then makes well-informed judgments based on the information it has acquired. One of the key advantages of deep learning compared to traditional ML methods is its ability to perform feature engineering automatically. In order to learn quickly, a deep learning system may analyse the data for related elements without being explicitly instructed. We used the MAGIC gamma telescope dataset, which has 19020 observations with ten characteristics (excluding the target). As a result, we proposed a deep learning-based strategy for identifying gamma rays.

The major contributions of this article are:

- We proposed an efficient classification model to identify high-energy gamma particles in Cherenkov Gamma Telescope data using deep learning techniques. The proposed classification model consists of several stages: pre-processing, exploratory data analysis, square root transformation, dataset segregation, deep learning classifier, and experimental validation.
- We presented an exploratory data analysis that involves analyzing the relationships between the target attribute 'class' and various features. Scatter plots and heat maps are used to visualize correlations and patterns. We further normalized the skewed distributions of attributes using square root transformation.
- We trained and evaluated three deep learning models: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (Bi-LSTM). Different activation functions (ReLU and Swish) are used in hidden layers to improve model performance. The models are evaluated using various metrics: accuracy, precision, recall, F1 score, and Area Under the ROC Curve (AUC).
- Our proposed model, LSTM+ReLU, achieved the highest accuracy of 88.71% with all attributes and 88.76% when the least correlated attribute is removed. The AUC value for the LSTM+ReLU model is 0.9377, indicating strong performance in distinguishing between classes.
- The proposed model outperforms existing methods, including logistic regression, linear discriminant analysis, and CART. This research study demonstrates that the proposed LSTM+ReLU model is effective in identifying gamma particles, with an accuracy of up to 88.76%. The model's AUC-ROC value further validates its efficacy in classifying gamma and hadron particles.

The research contributes to the field of gamma-ray astronomy by providing insights into the classification of high-energy gamma particles using deep learning techniques.

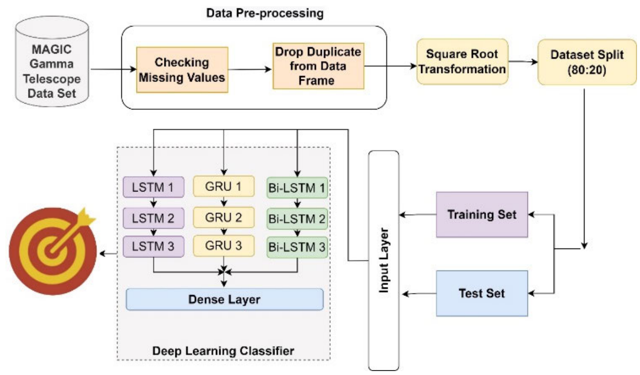


FIGURE 1. The block diagram of proposed classification model.

TABLE 1. Description of the dataset.

Feature	Description
fLength	The length of the major axis of an ellipse fitted to the particle image in millimetres.
fWidth	The length of the minor axis of an ellipse fitted to the particle image in millimetres.
fSize	The logarithm of the sum of the content of all pixels in the particle image.
fConc	The ratio of the sum of the two highest pixel values to fSize.
fConc1	The ratio of the highest pixel value to fSize.
fAsym	The distance in millimetres from the highest pixel to the center of the particle image, projected onto the major axis.
fM3Long	The third root of the third moment of the particle image along the major axis in millimetres.
fM3Trans	The third root of the third moment of the particle image along the minor axis in millimetres.
fAlpha	fAlpha represents the angle of the major axis of the ellipse fitted to the particle image with respect to the vector to the origin, measured in degrees.
fDist	It is the distance in millimetres from the origin to the centre of the ellipse.
class	The class variable categorizes the type of particle as either "g-gamma" (signal) or "h-hadron" (background).

III. METHODOLOGY

Figure 1 illustrates the block diagram of the proposed classification model. It is composed of several stages, including pre-processing, exploratory data analysis, square root transformation, dataset segregation, deep learning classifier, and experimental validation.

A. DATA

The MAGIC gamma telescope dataset, sourced from the UCI ML repository [26], is designed to simulate the registration of high-energy gamma particles using the IACT. The Cherenkov gamma telescope studies high-energy gamma rays by detecting the Cherenkov radiation emitted by charged particles generated in electromagnetic showers established by gamma rays. This dataset comprises 19020 instances, each with 10 attributes, and its description is provided in Table 1. The target variable 'class' has two values: 'g' (gamma) for signal and 'h' (hadron) for background, occurring with frequencies of 12332 and 6688, respectively.

The dataset contains information about the pulses left by incoming Cherenkov photons on a plane of photomultiplier

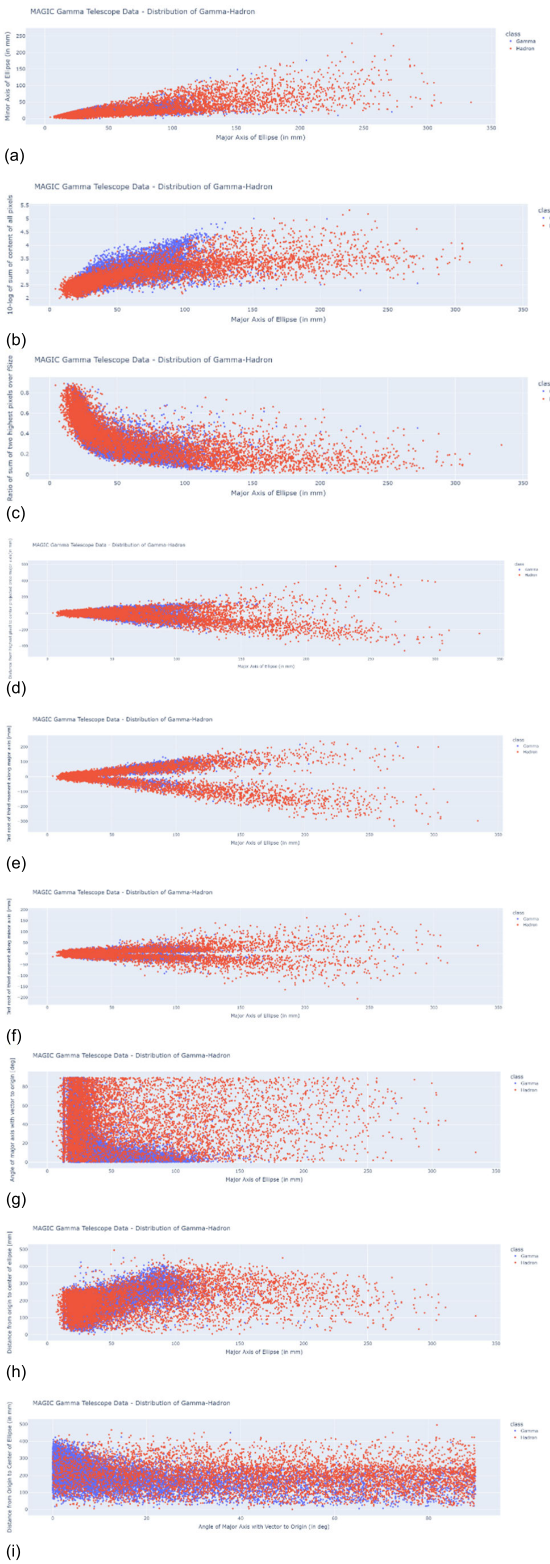


FIGURE 2. Scatter plot facet – target distribution (a) fLength Vs fWidth (b) fLength Vs fSize (c) fLength Vs fConc (d) fLength Vs fAsym (e) fLength Vs fM3Long (f) fLength Vs fM3Trans (g) fLength Vs fAlpha (h) fLength Vs fDist (i) fAlpha Vs fDist.

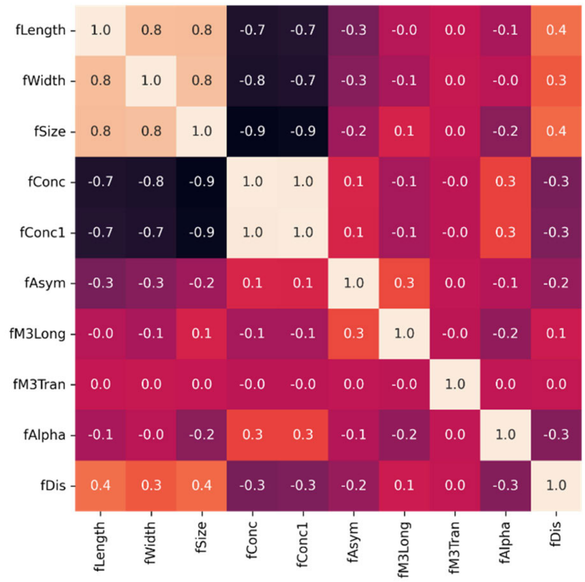


FIGURE 3. Heat map.

tubes, which is commonly referred to as the camera. This data allows for discriminating between signal and background by reconstructing the shower image, which represents the resulting pattern from the Cherenkov photons. The attribute information includes 11 continuous variables, such as the major and minor axis of the ellipse, size, concentration, asymmetry, and angle of the major axis with the vector to origin.

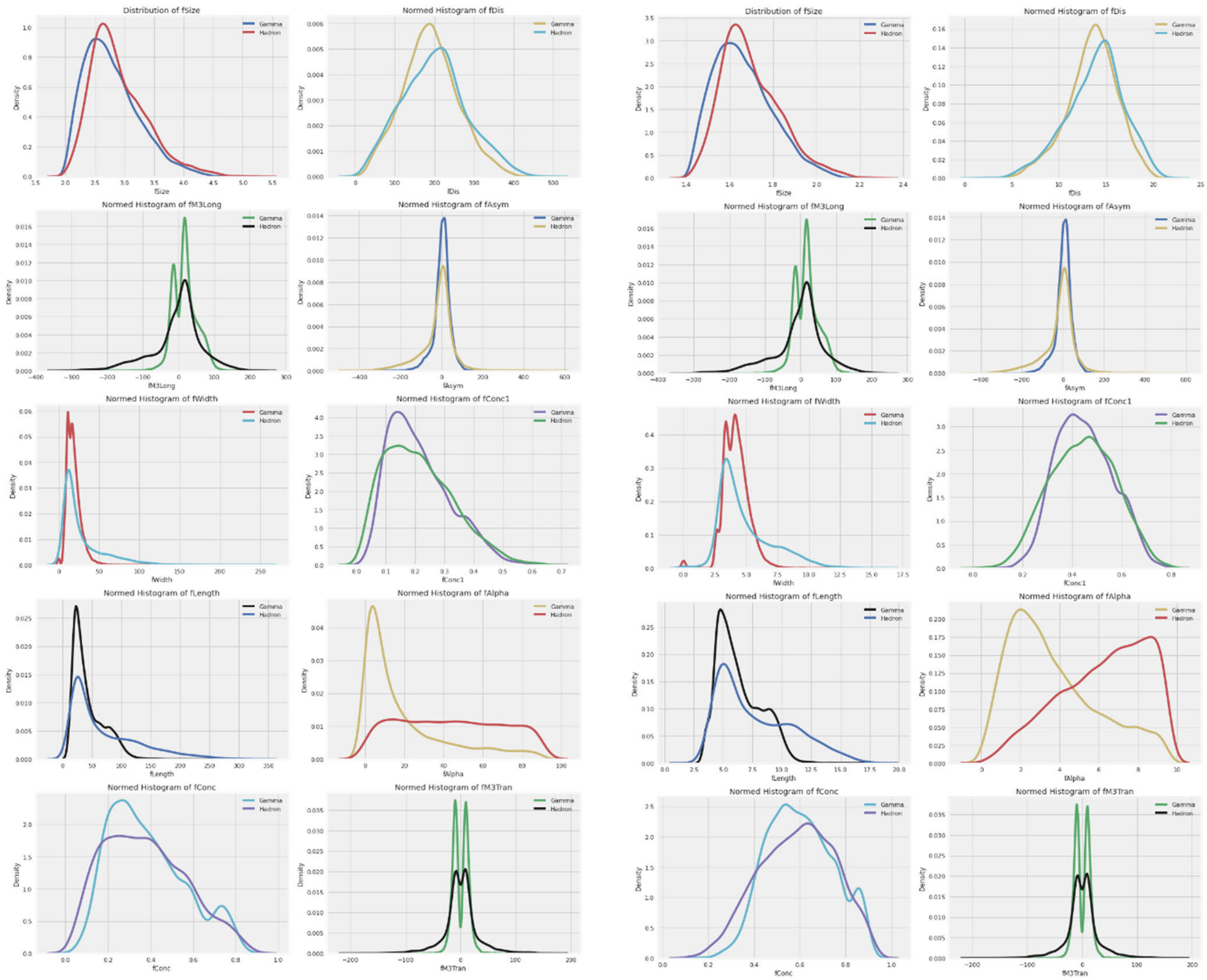
However, it is worth noting that the dataset has a limitation, as it does not provide information about the distribution of energy and zenith for the 19020 events.

B. DATA PRE-PROCESSING

Duplicate values and missing values have been identified and removed. There are 18905 distinct instances. The number of distinct gamma and hadron instances are 12,326 and 6579, respectively. Duplicate values can cause biased results and artificially inflate the performance of the model, while missing values can lead to inaccurate predictions or biased results if not handled properly. By removing duplicates and missing values, the dataset becomes more reliable and the model can learn patterns and make predictions more accurately.

C. EXPLORATORY DATA ANALYSIS

Figure 2 displays the scatter plot facet, highlighting the distribution of the target attribute ‘class’. Each subplot (a) to (h) demonstrates the distribution of ‘class’ and the correlation between various features, including fLength, fWidth, fSize, fConc, fAsym, fM3Long, fM3Trans, fAlpha, and fDist. Additionally, subplot 3. (i) illustrates the distribution of ‘class’ and the correlation between fAlpha and fDist. Once subjected to preprocessing, a shower image typically appears as a long cluster. When the telescope is pointed at a point source and the shower axis aligns with the telescope’s optical axis, the telescope’s long axis becomes perpendicular to the camera’s centre. In this scenario, an ellipse is defined by performing a principal component analysis in the camera plane, yielding a correlation axis. The unique properties of the ellipse (referred to as Hillas parameters) can be utilized as image attributes



(a) (b)

FIGURE 4. **(a) Original dataset distribution plot; (b) Square root transformation dataset distribution plot.**

for discrimination. Asymmetric energy deposits along the principal axis offer a means of differentiation. Additionally, other criteria such as the cluster's size on the image plane or the total number of depositions can be employed for discrimination. It is evident from the figure that the length parameter exhibits a stronger dependence on energy compared to the width parameter.

Figure 3 depicts the pre-processed dataset's heat map. According to the heat map, the 'fM3Tran' attribute has the lowest correlation with all other attributes. As a result, it is determined to conduct the two sorts of investigations. One includes all ten features, while the other excludes the ''fM3Tran' attribute and conducts the assessment on deep learning classifiers.

D. SQUARE ROOT TRANSFORMATION

A square root transformation can be preferable for normalising a skewed distribution. Data distributions can be described as symmetric (low skewness) or asymmetric (high skewness), and as heavy (high kurtosis) or light (low kurtosis) tails (normal distribution). It is possible for data to be positively skewed (with data skewed to the right) or negatively skewed (with data slanted (data pushed towards the left side) [27]. If the variable has right-skewed data, a square root transformation can be used to normalise it [28]. Figure 4 (a) shows the distribution plot of the original dataset with 18905 unique instances. It is observed that most of the features are skewed, and their values can be found in table 2. Except for 'fAsym', 'fM3Long', and 'fM3Tran' attributes, the other 7 attributes have been performed with square root transformation. The minimised skew and kurtosis values can be found in table 2. The same can be observed in figure 4 (b). It shows the distribution plot after square root transformation.

E. DEEP LEARNING

Long short-term memory (LSTM) and gated recurrent unit (GRU) belong to the category of recurrent neural networks (RNNs). Each layer in these networks takes its input,

TABLE 2. Magic gamma telescope dataset – skew and kurtosis values.

Attributes	Original Dataset		After Square Root Transform	
	skew	**kurtosis**	**skew**	**kurtosis**
fLength	2.022	5.031	1.137	1.062
fWidth	3.395	17.013	1.597	4.809
fSize	0.873	0.723	0.657	0.195
fConc	0.489	-0.517	-0.023	-0.572
fConc1	0.687	0.031	0.088	-0.428
fAsym	-1.038	8.231	-1.038	8.231
fM3Long	-1.130	4.717	-1.130	4.717
fM3Tran	0.124	8.676	0.124	8.676
fAlpha	0.857	-0.521	0.272	-1.125
fDis	0.229	-0.112	-0.391	0.171

multiplies it by a linear layer, and then adds the result to the hidden layer weights. This output is passed on to the next iteration of the network, creating a feedback loop characteristic of recurrent neural networks. In contrast to traditional feedforward neural networks, LSTMs include feedback connections [29]. GRUs require fewer training parameters, leading to reduced memory usage and faster execution compared to LSTMs. However, on larger datasets, LSTM tends to be more accurate [30]. Both LSTM and GRU models have an advantage over ordinary RNNs as they mitigate the problem of the vanishing gradient. Specifically, LSTM and GRU show better performance in terms of validation and prediction accuracy [31].

Additionally, a Bidirectional LSTM (Bi-LSTM) is a sequence processing model that comprises two LSTMs—one processing inputs in a forward direction and the other processing inputs in a backward direction [32]. In this study, LSTM, GRU, and Bi-LSTM networks were utilized for classification purposes.

1) ACTIVATION FUNCTION

With different NN models and datasets, activation functions behave differently [33]. The ReLU function (rectified linear activation function) is a piecewise linear function that returns the input value unmodified if the value is positive and returns zero otherwise. Two famous types of nonlinear activation functions are the sigmoid and the hyperbolic tangent. With both the sigmoid and tanh functions, saturation is a frequent problem. Therefore, large tanh and sigmoid values snap to 1.0, and small ones to -1.0 or 0.0. The Google Brain Team released a activation function called Swish, and it's as easy as $f(x) = x$ sigmoid (x). Their findings suggest that when it comes to deeper models, Swish is superior to ReLU. So, it is decided to validate the deep learning approaches with Swish and ReLU activation functions.

The ReLU activation function [34] can be defined as,

$$f(x) = \begin{Bmatrix} x & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{Bmatrix} \tag{1}$$

where x is the input to a neuron.

The swish activation function [35] can be expressed as,

$$\text{swish}(x) = x.\text{sigmoid}(\beta x) = \frac{x}{1 + e^{-\beta x}} \tag{2}$$

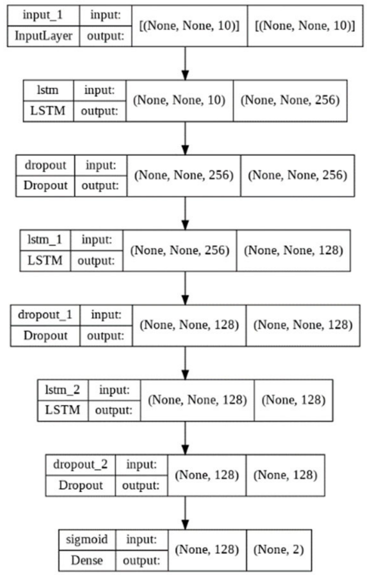


FIGURE 5. Proposed LSTM+ReLU & LSTM+swish model.

where β is either constant or a trainable parameter depending on the model.

2) LSTM

Figure 5 depicts the proposed LSTM model. It is comprised of three hidden layers, the first of which is an LSTM base layer, followed by a standard feedforward output layer. The activation functions "ReLU" and "swish" were used in hidden layers. Because the dataset is binary in nature, the 'sigmoid' activation function was used in the dense layer. The "sparse_categorical_crossentropy" as loss function and "Adam" as optimizer have been employed in the model.

The equations (3) to (8) show in concise form the forward pass of an LSTM cell with a forget gate [36]. The lowercase variables represent vectors.

$$f_t = \sigma_g (W_f x_t + U_f h_{t-1} + b_f) \tag{3}$$
$$i_t = \sigma_g (W_i x_t + U_i h_{t-1} + b_i) \tag{4}$$
$$o_t = \sigma_g (W_o x_t + U_o h_{t-1} + b_o) \tag{5}$$
$$\tilde{C}_t = \sigma_c (W_c x_t + U_c h_{t-1} + b_c) \tag{6}$$
$$C_t = f_t \odot c_{t-1} + i_t \odot \tilde{C}_t \tag{7}$$
$$h_t = o_t \odot \sigma_h (C_t) \tag{8}$$

3) GRU

The GRU's operation is analogous to that of an LSTM equipped with a forget gate, however it has fewer parameters due to the absence of an output gate [37]. Figure 6 depicts the proposed GRU+ReLU model. It is comprised of three hidden layers, the first of which is a base layer, followed by a standard feedforward output layer. Because the dataset is binary in nature, the 'sigmoid' activation function was used in the dense layer. The "sparse_categorical_crossentropy" as

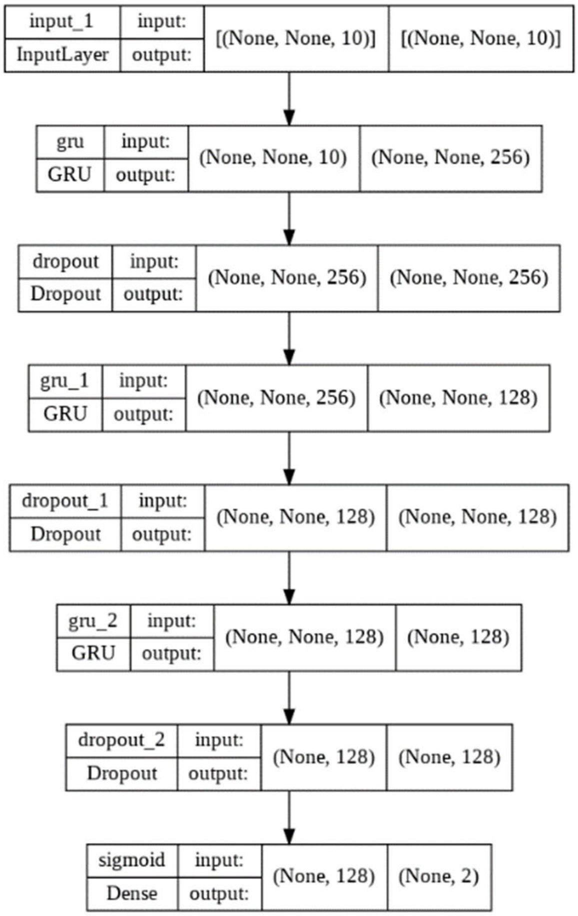


FIGURE 6. Proposed GRU+ReLU model.

loss function and "Adam" as optimizer have been employed in the model. Equations (9) and (10), respectively, express the update gate vector and reset gate vector of GRU.

$$\mathrm{z_t} = \sigma_g\left(\mathrm{U_z h_{t-1}} + \mathrm{b_z}\right) \tag{9}$$

$$\mathrm{r_t} = \sigma_g\left(\mathrm{U_r h_{t-1}} + \mathrm{b_r}\right) \tag{10}$$

4) BI-LSTM

Bi-LSTM can learn long-term dependencies without keeping redundant background information [38]. Figure 7 depicts the proposed Bi-LSTM+ReLU model. It is comprised of three hidden layers, the first of which is a base layer, followed by a standard output layer. Because the dataset is binary, the sigmoid activation function has been used in the dense layer. In the model, the loss function "sparse categorical crossentropy" and the optimizer "Adam" were used.

F. RESULTS AND DISCUSSION

The preprocessed dataset is consisting of 18905 distinct instances. The dataset has been divided into 15124 and 3781 occurrences as training and test sets, respectively. The testing set was used to evaluate the model's performance. The identification of Gamma particles has been performed

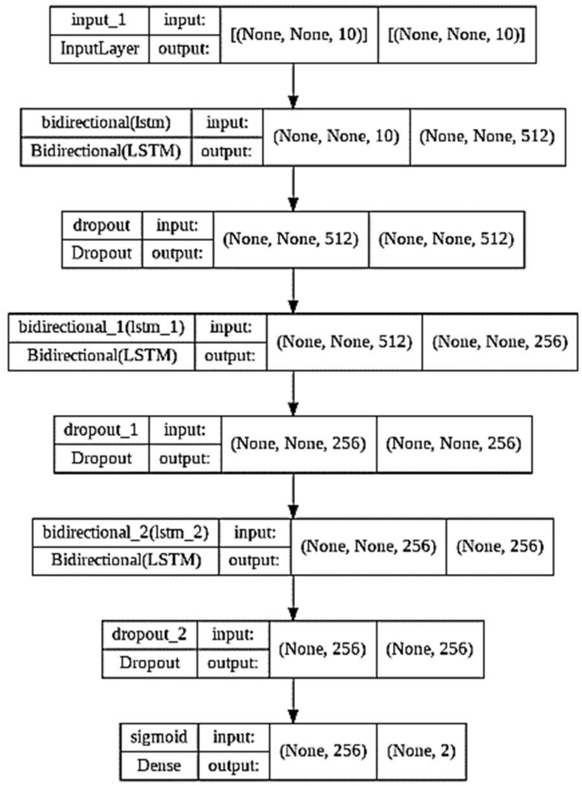


FIGURE 7. Proposed Bi-LSTM+ReLU model.

with two cases as shown in Table 3. In first case, all ten attributes have been used and classification accuracy has been validated with the deep learning models LSTM+ReLU, GRU+ReLU, Bi-LSTM+ReLU, and LSTM+Swish. In Second case, As per the correlation matrix, the 'fM3Tran' attribute has the lowest correlation with all other attributes. So, to test the accuracy of classification, this feature has been discarded, and the remaining 9 attributes with LSTM+Swish and LSTM+ReLU have been validated. Figure 8 shows the loss function plot of (a) LSTM+ReLU (b) GRU+ReLU (c) Bi-LSTM + ReLU (d) LSTM+Swish (e) LSTM+Swish with 9 attributes (f) LSTM+ReLU with 9 attributes models. A loss function is a function that compares the predicted and target output values. The loss function is a mechanism for determining how effectively a classification model models the dataset. A 'Sparse categorical cross entropy' loss function, a "sigmoid" activation function, and an "ADAM" optimizer with 50 epochs have been performed for validation. Each iteration of the model's training with all of the available data is called an epoch. The model's performance stabilizes around the 50th epoch, indicating a convergence point where further training epochs may not yield significant improvements and might lead to overfitting. Figure 9 shows the confusion matrix of the classification model. In confusion matrix, the gamma (signal) is represented as 1 and hadron (background) is represented as 0. Figure 10 shows the receiver operating characteristic (ROC) curve and area under the ROC Curve (AUC) values of each model. Table 3 lists the classification metrics such

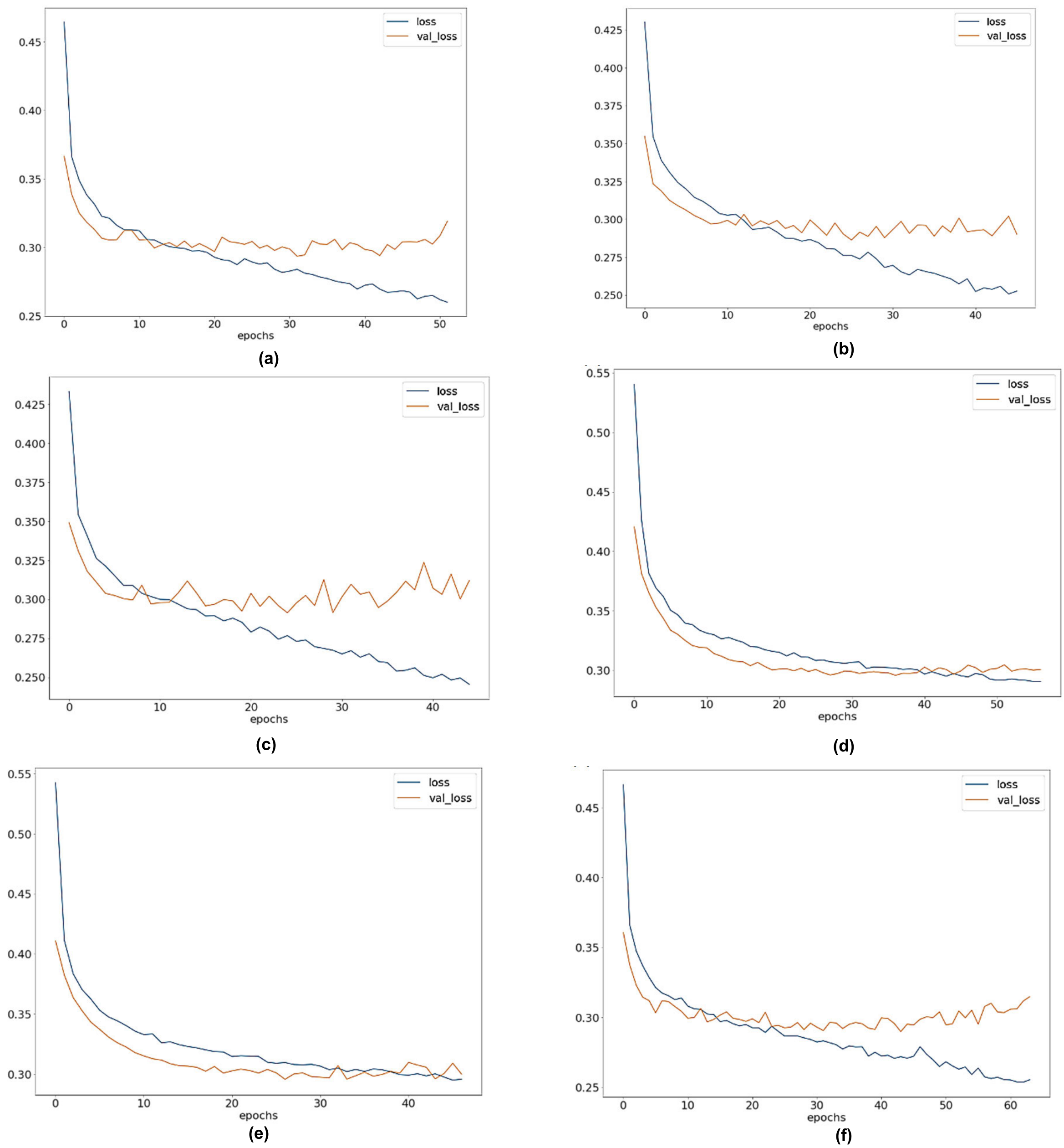


FIGURE 8. Loss function plot (a) LSTM+ReLU (b) GRU+ReLU (c) Bi-LSTM + ReLU (d) LSTM+Swish (e) LSTM+Swish with 9 attributes (f) LSTM+ReLU with 9 attributes.

as accuracy, precision, recall, F1 score, and AUC of each model [39]. The training set accuracy of the LSTM+ReLU model has been obtained as 88.69%. It has been seen that the LSTM+ReLU model has a better accuracy of 88.71% with all the attributes for test set, and if the attribute with the least correlation, fM3Tran, is taken out of the model, the accuracy is 88.76%.

The performance parameters were expressed in equations from (11) to (14).

$$\text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) \quad (11)$$

$$\text{Precision}(p) = \text{TP}/(\text{TP} + \text{FP}) \quad (12)$$

$$\text{Recall}(r) = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

Here, TP stands for true positive, TN stands for true negative, FP stands for false positive, and FN stands for false negative.

$$\text{f1 Score} = (2pr)/(p + r) \quad (14)$$

The findings [40] of the area under the ROC curve (AUC) were regarded as excellent for AUC values ranging from 0.9 to 1. The LSTM+ReLU model has an AUC value of 0.9377. For the selected problem statement, the

TABLE 3. **MAGIC gamma telescope data set – test set.**

Model	Accuracy		Precision		Recall		F1 Score		AUC	
	Gamma	Hadron	Gamma	Hadron	Gamma	Hadron	Gamma	Hadron	Gamma (1)	Hadron (0)
No. of Attributes = 10										
LSTM+ReLU	0.8871	0.8871	0.8788	0.9073	0.9585	0.7543	0.9169	0.8238	0.9377	0.9376
GRU+ReLU	0.8836	0.8836	0.8831	0.8849	0.9463	0.7672	0.9136	0.8219	0.9395	0.9394
Bi-LSTM + ReLU	0.885	0.885	0.877	0.9044	0.9573	0.7506	0.9154	0.8203	0.9396	0.9397
LSTM+Swish	0.876	0.876	0.8721	0.8854	0.9483	0.7415	0.9086	0.8071	0.9320	0.9320
No. of Attributes = 9										
LSTM+Swish	0.8736	0.8736	0.8691	0.8844	0.9483	0.7347	0.907	0.8026	0.9303	0.9302
LSTM+ReLU	0.8876	0.8876	0.8783	0.9104	0.9601	0.7528	0.9174	0.8242	0.9417	0.9417

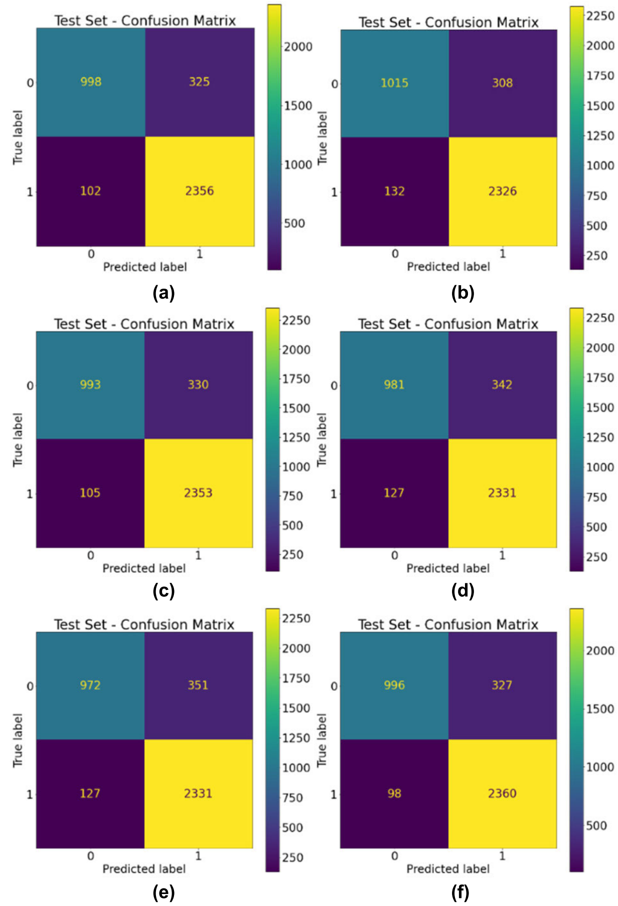


FIGURE 9. **Confusion matrix (a) LSTM+ReLU (b) GRU+ReLU (c) Bi-LSTM + ReLU (d) LSTM+Swish (e) LSTM+Swish with 9 attributes (f) LSTM+ReLU with 9 attributes.**

TABLE 4. **Comparison with existing methods.**

	Model	Accuracy in %
Emmanuel Dadzie and Kelvin Kwakye (2021) [41]	Logistic regression	76.04
	Linear discriminant analysis	76.06
	Classification And Regression Trees (CART)	79.44
Our proposed Approach	LSTM+ReLU	88.71 with 10 attributes
	LSTM+ReLU	88.76 with 9 attributes

LSTM+ReLU model performs better than other models such as GRU+ReLU, Bi-LSTM+ReLU, and LSTM+Swish.

The integration of Long Short-Term Memory (LSTM) networks with Rectified Linear Unit (ReLU) activation functions aims to leverage the strengths of both architectures. LSTMs are proficient in capturing long-term dependencies, while ReLU offers efficient and non-linear transformations, enabling better representation learning. By integrating LSTM layers with ReLU activations, the model combines the memory-retaining capabilities of LSTMs with the non-linear transformation and computational efficiency of ReLU units. This hybrid architecture enables the model to capture intricate patterns and dependencies within the data while maintaining computational efficiency. The hybrid architecture facilitates improved feature learning and representation, enabling the model to capture complex relationships and patterns within the data more effectively. The combination of LSTM and ReLU helps alleviate gradient-related issues, such as vanishing and exploding gradients, by leveraging the stability and non-linearity introduced by ReLU activations. ReLU activations contribute to computational efficiency by accelerating the training process and reducing computational overhead, allowing the model to process and analyze data more efficiently. Experimental evaluations and performance metrics demonstrate that the hybridization of LSTM and ReLU architectures consistently yields superior results compared to other combinations. Table 4 presents a comparison between the proposed approach and the existing methods. In a recent study by Emmanuel Dadzie and Kelvin Kwakye [41], an accuracy of 79.44% was achieved using the CART ML method. In contrast, our results demonstrate a significantly higher accuracy of 88.76% with 9 attributes and 88.71% with 10 attributes, indicating the superiority of our proposed approach.

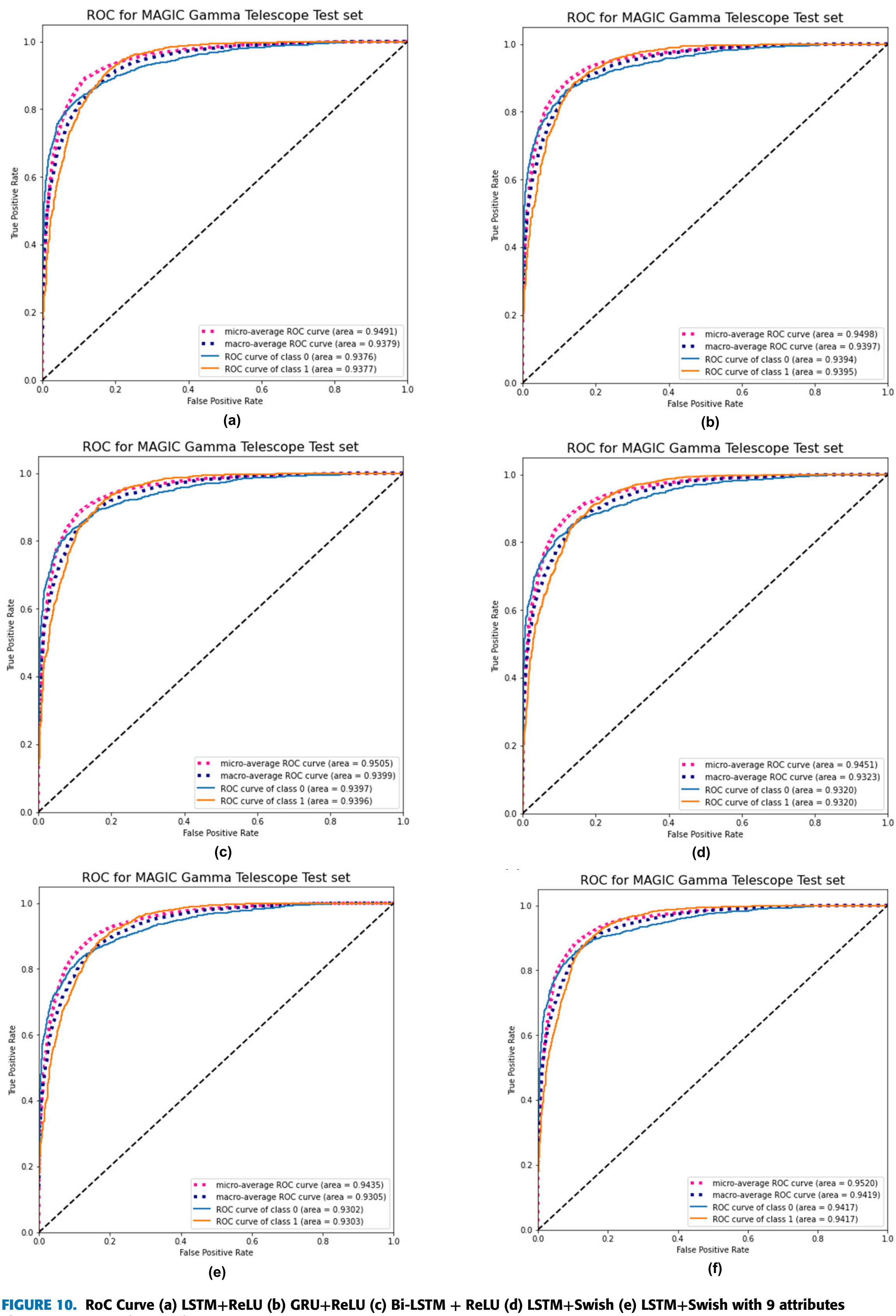


FIGURE 10. **RoC Curve (a) LSTM+ReLU (b) GRU+ReLU (c) Bi-LSTM + ReLU (d) LSTM+Swish (e) LSTM+Swish with 9 attributes (f) LSTM+ReLU with 9 attributes.**

IV. CONCLUSION

Utilizing deep learning algorithms, this study distinguishes between simulated gamma and hadron events reconstructed by the MAGIC Cherenkov Telescope. Through rigorous validation, we evaluated various deep learning models, including LSTM+ReLU, GRU+ReLU, Bi-LSTM+ReLU, and LSTM+Swish, using the MAGIC Gamma Telescope Data Set. To address data distribution issues, a square root transformation was implemented, focusing on skewness and kurtosis adjustments.

Our findings distinctly highlight the LSTM+ReLU model's prowess, achieving an impressive classification accuracy of 88.71% when considering all attributes. Notably, even upon excluding the least correlated attribute, fM3Tran, the accuracy remains virtually unchanged at 88.76%. To further validate the model's discriminative capability, we employed the AUC-ROC metric, where the LSTM+ReLU model prominently recorded an AUC value of 0.9377. This robust performance solidifies its effectiveness in accurately identifying gamma rays. The limitation of this research is the foundational dataset from the UCI ML repository lacks crucial energy and zenith distribution information, which could affect the model's accuracy and generalizability to real-world scenarios. The absence of certain features might limit the model's ability to capture the full complexity of gamma-ray events.

ACKNOWLEDGMENT

The authors extend their appreciation to King Saud University for funding this research through Researchers Supporting Project Number (RSPD2024R890), King Saud University, Riyadh, Saudi Arabia.

REFERENCES

[1] T. C. Weekes, M. F. Cawley, D. J. Fegan, K. G. Gibbs, A. M. Hillas, P. W. Kowk, R. C. Lamb, D. A. Lewis, D. Macomb, N. A. Porter, P. T. Reynolds, and G. Vacanti, "Observation of TeV gamma rays from the crab nebula using the atmospheric cerenkov imaging technique," *Astrophysical J.*, vol. 342, p. 379, Jul. 1989, doi: 10.1086/167599.

[2] D. Caprioli, "An original mechanism for the acceleration of ultra-high-energy cosmic rays," *Nucl. Part. Phys. Proc.*, vols. 297–299, pp. 226–233, Apr. 2018, doi: 10.1016/j.nuclphysbps.2018.07.032.

[3] D. Gaggero and M. Valli, "Impact of cosmic-ray physics on dark matter indirect searches," *Adv. High Energy Phys.*, vol. 2018, pp. 1–23, Dec. 2018, doi: 10.1155/2018/3010514.

[4] D. Horns and A. Jacholkowska, "Gamma rays as probes of the universe," *Comp. Rendus Phys.*, vol. 17, no. 6, pp. 632–648, Jun. 2016, doi: 10.1016/j.crhy.2016.04.006.

[5] F. Arneodo, A. Di Giovanni, and P. Marpu, "A review of requirements for gamma radiation detection in space using CubeSats," *Appl. Sci.*, vol. 11, no. 6, p. 2659, Mar. 2021, doi: 10.3390/app11062659.

[6] M. de Naurois and D. Mazin, "Ground-based detectors in very-high-energy gamma-ray astronomy," *Comp. Rendus Phys.*, vol. 16, nos. 6–7, pp. 610–627, Aug. 2015, doi: 10.1016/j.crhy.2015.08.011.

[7] F. M. Rieger, E. de Oña-Wilhelmi, and F. A. Aharonian, "TeV astronomy," *Frontiers Phys.*, vol. 8, no. 6, pp. 714–747, Dec. 2013, doi: 10.1007/s11467-013-0344-6.

[8] C. Arcaro, D. Corti, A. De Angelis, M. Doro, C. Manea, M. Mariotti, R. Rando, I. Reichardt, and D. Tescaro, "Studies on a silicon-photomultiplier-based camera for imaging atmospheric Cherenkov telescopes," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 876, pp. 26–30, Dec. 2017, doi: 10.1016/j.nima.2016.12.055.

[9] R. Cristin, B. S. Kumar, C. Priya, and K. Karthick, "Deep neural network based rider-cuckoo search algorithm for plant disease detection," *Artif. Intell. Rev.*, vol. 53, no. 7, pp. 4993–5018, Oct. 2020, doi: 10.1007/s10462-020-09813-w.

[10] K. Sekar, K. Kanagarathinam, S. Subramanian, E. Venugopal, and C. Udayakumar, "An improved power quality disturbance detection using deep learning approach," *Math. Problems Eng.*, vol. 2022, pp. 1–12, May 2022, doi: 10.1155/2022/7020979.

[11] D.-E. Choe, H.-C. Kim, and M.-H. Kim, "Sequence-based modeling of deep learning with LSTM and GRU networks for structural damage detection of floating offshore wind turbine blades," *Renew. Energy*, vol. 174, pp. 218–235, Aug. 2021, doi: 10.1016/j.renene.2021.04.025.

[12] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting," *IEEE Access*, vol. 8, pp. 143759–143768, 2020, doi: 10.1109/ACCESS.2020.3009537.

[13] H. Liu and G. Cao, "Deep learning video analytics through online learning based edge computing," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8193–8204, Oct. 2022, doi: 10.1109/TWC.2022.3164598.

[14] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, and T. Sainath, "Deep learning for audio signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 2, pp. 206–219, May 2019, doi: 10.1109/JSTSP.2019.2908700.

[15] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text data augmentation for deep learning," *J. Big Data*, vol. 8, no. 1, p. 101, Dec. 2021, doi: 10.1186/s40537-021-00492-0.

[16] X.-W. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014, doi: 10.1109/ACCESS.2014.2325029.

[17] C. Zhang, G. Hu, F. Luo, Y. Xiang, G. Ding, C. Chu, J. Zeng, R. Ze, and Q. Xiang, "Identification of SNM based on low-resolution gamma-ray characteristics and neural network," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 927, pp. 155–160, May 2019, doi: 10.1016/j.nima.2019.02.023.

[18] M. Sharma, J. Nayak, M. K. Koul, S. Bose, and A. Mitra, "Gamma/hadron segregation for a ground based imaging atmospheric Cherenkov telescope using machine learning methods: Random forest leads," *Res. Astron. Astrophys.*, vol. 14, no. 11, pp. 1491–1503, Nov. 2014, doi: 10.1088/1674-4527/14/11/012.

[19] R. K. Bock, A. Chilingarian, M. Gaug, F. Hakl, T. Hengstebeck, M. Jirina, J. Klaschka, E. Kotrč, P. Savický, S. Towers, A. Vaiciulis, and W. Wittek, "Methods for multidimensional event classification: A case study using images from a Cherenkov gamma-ray telescope," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 516, nos. 2–3, pp. 511–528, Jan. 2004, doi: 10.1016/j.nima.2003.08.157.

[20] H. Hirashima, T. Ono, M. Nakamura, Y. Miyabe, N. Mukumoto, H. Iramina, and T. Mizowaki, "Improvement of prediction and classification performance for gamma passing rate by using plan complexity and dosiomics features," *Radiotherapy Oncol.*, vol. 153, pp. 250–257, Dec. 2020, doi: 10.1016/j.radonc.2020.07.031.

[21] S. Ohm, C. van Eldik, and K. Egberts, "γ/hadron separation in very-high-energy γ-ray astronomy using a multivariate analysis method," *Astroparticle Phys.*, vol. 31, no. 5, pp. 383–391, Jun. 2009, doi: 10.1016/j.astropartphys.2009.04.001.

[22] M. Krause, E. Pueschel, and G. Maier, "Improved γ/hadron separation for the detection of faint γ-ray sources using boosted decision trees," *Astroparticle Phys.*, vol. 89, pp. 1–9, Mar. 2017, doi: 10.1016/j.astropartphys.2017.01.004.

[23] A. Brill, Q. Feng, T. B. Humensky, B. Kim, D. Nieto, and T. Miener, "Investigating a deep learning method to analyze images from multiple gamma-ray telescopes," in *Proc. New York Scientific Data Summit (NYSDS)*, 2019, pp. 1–4, doi: 10.1109/NYSDS.2019.8909697.

[24] D. N. Castaño, A. Brill, Q. Feng, T. B. Humensky, B. Kim, T. Miener, R. Mukherjee, and J. Sevilla, "CTLearn: Deep learning for gamma-ray astronomy," in *Proc. 36th Int. Cosmic Ray Conf.*, Jul. 2019, pp. 1–8.

[25] R. Conceição, B. S. González, A. Guillén, M. Pimenta, and B. Tomé, "Muon identification in a compact single-layered water Cherenkov detector and gamma/hadron discrimination using machine learning techniques," *Eur. Phys. J. C*, vol. 81, no. 6, p. 542, Jun. 2021, doi: 10.1140/epjc/s10052-021-09312-4.

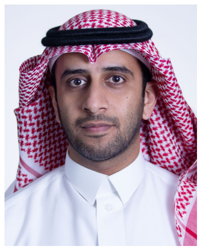
[26] R. Bock, "MAGIC gamma telescope," UCI Mach. Learn. Repository, 2007. [Online]. Available: https://doi.org/10.24432/C52C8B

[27] R. Bono, J. Arnau, R. Alarcón, and M. J. Blanca, ‘‘Bias, precision, and accuracy of skewness and kurtosis estimators for frequently used continuous distributions,’’ *Symmetry*, vol. 12, no. 1, p. 19, Dec. 2019, doi: 10.3390/sym12010019.
[28] J. H. McDonald, *Handbook of Biological Statistics*, 3rd ed. Baltimore, MD, USA: Sparky House, 2014.
[29] Z. Zhang, H. Cheng, and T. Yang, ‘‘A recurrent neural network framework for flexible and adaptive decision making based on sequence learning,’’ *PLOS Comput. Biol.*, vol. 16, no. 11, Nov. 2020, Art. no. e1008342, doi: 10.1371/journal.pcbi.1008342.
[30] C. Wang, W. Du, Z. Zhu, and Z. Yue, ‘‘The real-time big data processing method based on LSTM or GRU for the smart job shop production process,’’ *J. Algorithms Comput. Technol.*, vol. 14, pp. 1–9, Jan. 2020, doi: 10.1177/1748302620962390.
[31] S.-H. Noh, ‘‘Analysis of gradient vanishing of RNNs and performance comparison,’’ *Information*, vol. 12, no. 11, p. 442, Oct. 2021, doi: 10.3390/info12110442.
[32] R. L. Abduljabbar, H. Dia, and P.-W. Tsai, ‘‘Unidirectional and bidirectional LSTM models for short-term traffic prediction,’’ *J. Adv. Transp.*, vol. 2021, pp. 1–16, Mar. 2021, doi: 10.1155/2021/5589075.
[33] I. Jahan, M. F. Ahmed, M. O. Ali, and Y. M. Jang, ‘‘Self-gated rectified linear unit for performance improvement of deep neural networks,’’ *ICT Exp.*, vol. 9, no. 3, pp. 320–325, Jun. 2023, doi: 10.1016/j.icte.2021.12.012.
[34] C. Banerjee, T. Mukherjee, and E. Pasiliao, ‘‘The multi-phase ReLU activation function,’’ in *Proc. ACM Southeast Conf.* New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 239–242, doi: 10.1145/3374135.3385313.
[35] G. C. Tripathi, M. Rawat, and K. Rawat, ‘‘Swish activation based deep neural network predistorter for RF-PA,’’ in *Proc. TENCON IEEE Region Conf. (TENCON)*, Oct. 2019, pp. 1239–1242, doi: 10.1109/TENCON.2019.8929500.
[36] H. Mayer, F. Gomez, D. Wierstra, I. Nagy, A. Knoll, and J. Schmidhuber, ‘‘A system for robotic heart surgery that learns to tie knots using recurrent neural networks,’’ in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2006, pp. 543–548, doi: 10.1109/IROS.2006.282190.
[37] L. Bi, G. Hu, M. M. Raza, Y. Kandel, L. Leandro, and D. Mueller, ‘‘A gated recurrent units (GRU)-based model for early detection of soybean sudden death syndrome through time-series satellite imagery,’’ *Remote Sens.*, vol. 12, no. 21, p. 3621, Nov. 2020, doi: 10.3390/rs12213621.
[38] Z. Xueqing, Z. Zhansong, and Z. Chaomo, ‘‘Bi-LSTM deep neural network reservoir classification model based on the innovative input of logging curve response sequences,’’ *IEEE Access*, vol. 9, pp. 19902–19915, 2021, doi: 10.1109/ACCESS.2021.3053289.
[39] M. Fatourechi, R. K. Ward, S. G. Mason, J. Huggins, A. Schlögl, and G. E. Birch, ‘‘Comparison of evaluation metrics in classification applications with imbalanced datasets,’’ in *Proc. 7th Int. Conf. Mach. Learn. Appl.*, Dec. 2008, pp. 777–782, doi: 10.1109/ICMLA.2008.34.
[40] K. Hajian-Tilaki, ‘‘Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation,’’ *Caspian J. Intern. Med.*, vol. 4, no. 2, pp. 627–635, 2013.
[41] E. Dadzie and K. Kwakye, ‘‘Developing a machine learning algorithm-based classification models for the detection of high-energy gamma particles,’’ 2021, *arXiv:2111.09496*.

K. KARTHICK (Member, IEEE) received the B.E. degree in electrical and electronics engineering from Periyar University, Salem, India, and the M.E. degree in power electronics and drives and the Ph.D. degree in electrical engineering from Anna University, Chennai, India. He has over 18 years of teaching experience. He is currently an Associate Professor with the Department of Electrical and Electronics Engineering, GMR Institute of Technology, Rajam, India. His research interests include data analytics, machine learning, text detection and recognition, image processing, and electrical drives. He was a member of ISTE.

S. AKILA AGNES received the bachelor’s and master’s degrees in information technology from Anna University, India, and the Ph.D. degree in computer science and engineering from the Karunya Institute of Technology and Sciences, India, in 2022. She is currently a Senior Assistant Professor with the GMR Institute of Technology, Rajam, Andhra Pradesh. Her research interests include computer vision, machine learning, deep learning, and medical image processing.

S. SENDIL KUMAR received the B.E. degree in electrical and electronics engineering from the Adhiyamaan College of Engineering, Hosur, in 2000, the M.Tech. degree in energy system engineering from VIT University, Vellore, in 2004, and the Ph.D. degree in power system engineering from Anna University, Chennai, in 2012. He is currently a Professor with the S. A. Engineering College, Chennai. His research interest includes signal processing techniques applied to power system engineering.



SULTAN ALFARHOOD received the Ph.D. degree in computer science from the University of Arkansas. He is currently an Assistant Professor with the Department of Computer Science, King Saud University (KSU). Since joining KSU, in 2007, he has made several contributions to the field of computer science through his research and publications. His work includes proposing innovative approaches and techniques to enhance the accuracy and effectiveness of these systems. His recent publications have focused on using deep learning and machine learning techniques to address challenges in these domains. His research continues to make significant contributions to the field of computer science and machine learning. His work has been published in several high-impact journals and conferences. His research interests include machine learning, recommender systems, linked open data, text mining, and ML-based IoT systems.

MEJDL SAFRAN received the bachelor’s degree in computer science from King Saud University, in 2007, and the master’s and Ph.D. degrees in computer science from Southern Illinois University Carbondale, in 2013 and 2018, respectively. His Ph.D. dissertation was on developing efficient learning-based recommendation algorithms for top-N tasks and top-N workers in large-scale crowdsourcing systems. He has been an AI Consultant for several national and international agencies, since 2018. He is a passionate Researcher and an Educator in the field of artificial intelligence, with a focus on deep learning and its applications in various domains. He is currently an Assistant Professor of computer science with King Saud University, where he has been a Faculty Member, since 2008. He has been leading grant projects in the fields of AI in medical imaging and AI in smart farming. He has published more than 20 articles in peer-reviewed journals and conference proceedings. His current research interests include developing novel deep learning methods for image processing, pattern recognition, natural language processing, and predictive analytics, as well as modeling and analyzing user behavior and interest in online platforms.